The Structure and Function of Glycoforms

Thesis

UNRESTRICTED

The Structure and Function of Glycoforms

Pauline Mary Rudd B.Sc.

This thesis is offered for the Degree of Doctor of Philosophy (PhD) in the Department of Chemistry, The Open University, Milton Keynes
Submitted: May 17th 1995

This work was carried out in the Glycobiology Institute, The Department of Biochemistry, University of Oxford
Director: Professor R.A. Dwek

Date of submission: 17 May 1995
Date of award: 1 September 1995

This thesis is dedicated to

Martin Nicolas, Simon Christopher, Hilary Claire & Jonathan Richard

whose integrity and generosity, perception and endurance, indomitable courage and joy, friendship and resourcefulness constantly inspires and renews me.

to my parents,
Stanley Norman and Clarice Ruma Goldsmith

who taught me their skills, surrounded me with opportunities and supported my dreams with unconditional love,

and to
Colin Richard

who for more than 30 years has shared with me his vocation and his aspirations, and sustained me with his insight, wisdom and steadfast love.

It is also dedicated to all those who have taught me and freely shared with me their love of learning and hard won knowledge, and who, through commitment to their own search for truth, have given light to me.
Without the infinite richness of my heritage this thesis could not have been written.

.....

'I think I could', whispered Tinkerbell faintly, 'if children believed in fairies'.
From across the world came the sound of distant clapping and her fading light grew stronger.
'That', said Peter, 'is what comes of being a fairy.'
It is also what comes of being believed in.

Taken from J.M. Barrie
Peter Pan

.....

To everything there is a season, and a time for every purpose under the heaven:
A time to be born, and a time to die: a time to sow and a time to reap;
A time to kill and a time to heal; a time to break down and a time to build-up;
A time to mourn and a time to dance;
A time to cast away stones and a time to gather stones together;
A time to embrace and a time to refrain from embracing;
A time to seek and a time to lose; a time to keep and a time to cast away;
A time to rend and a time to sew;
A time to keep silence and a time to speak;
A time to love and a time to hate; a time for war and a time for peace.

Ecc. 3. 1

In the end neither you nor I lead the dance,
You and I do not need to know the way,
In the end all is harvest.

Anon.

Acknowledgement

My interest in sugars first began when I was at school in Bournemouth and met Dr. Peter Parr, who at that time was using his kitchen to prepare glucose-1-phosphate by incubating inorganic phosphate with phosphorylase which he extracted from potato juice. Over the next ten years together we extracted, purified and often crystallised some 30 sugar phosphates and related compounds, which we marketed under the name of Wessex Biochemicals (later Sigma London). In the course of this enterprise our long-suffering families tolerated our collections of raw materials which ranged from brewer's yeast and raw mussels to the *Sedum Spectabile* and Autumn crocus plants which were growing (presumably originally for decoration) in our front gardens. At moments of crisis they could be depended upon to peel, chip and mince potatoes, homogenise plants and mix sludges in our indispensable side paddle washing machine. When our sales increased it was my father who designed purpose built laboratories to our exacting requirements so that we could expand and develop. During that time under Peter's guidance I learned a great deal of science, felt constantly challenged and above all came to learn the importance of commitment and integrity. Later, during the years when the children were small, and I exchanged the side paddle washing machine for a much less versatile automatic model, many people generously kept me in touch with the world of Science. Sigma Chemical Company, with the bemused agreement of the Parochial Church Council of St. Nicolas, North Stoneham, built me a laboratory at home, and Mr. George Moss, the director of Commercial and Forensic Laboratories in Reading invariably found me interesting short projects whenever I could persuade someone to baby-sit. We moved to Milton Keynes as Stantonbury Campus opened, and there Mr. Michael Bendry, who was head of Chemistry, encouraged me to think at a level far beyond that which my job as an A-level science technician required, inspiring me with his clarity of thought and his gifts as a teacher. I spent one day a week at the Open University working on a PhD project, and I am eternally grateful to those members of the Chemistry Department who made this possible because they also worked until midnight each Monday to supervise me.

In Oxford I was able to return to a field related to the one which had originally inspired me when Professor Raymond Dwek took me into his research group and gave me the opportunity to learn from scratch some of the basic technology and science needed to work in Glycobiology. Many people patiently helped me to acquire some skills, especially Dr. David Ashford and Dr. Thomas Rademacher. More recently I owe a great deal to many colleagues in the Glycobiology Institute, in particular Dr. David Harvey, who has taught me some of the intricacies of mass spectrometry and enabled me to interpret the data, and to Dr. Mark Wormald who has spent many hours with me, modelling the glycoproteins I have been studying. I would also like to thank Mr. Brian Matthews who has analysed my samples by gas chromatography and always been available to solve technical problems with equipment. Mr. Geoffrey Guile has been responsible for maintaining and developing the HPLC systems; without his skill much of the chromatography data in this thesis would not have been collected. Oxford GlycoSystems have supported the sequencing facility and have invariably responded to pleas for help with the state-of-the art sequencing equipment in the Glycobiology Institute. In addition, I would like to thank Mrs. Judith Manley and the photographic department in Nuclear Physics who contributed their skills to the art work in this thesis.

I have been fortunate to have had the opportunity to collaborate with people from whom I have learned a great deal: especially Professor Ghislain Opdenakker of the Rega Institute, who has constantly supported and encouraged me for many years, enabling me to think more widely about the projects I have been working on. I have also been fortunate to have collaborated with Dr. Neil Barclay and Dr. Simon Davis in the William Dunn School, Oxford with whom I have had many helpful discussions. I would like to thank Professor Baruch Blumberg, who has taken an interest in all I have hoped to achieve, and with whom I have had many important discussions about the philosophy of science. I particularly appreciated the opportunity which I was given to spend two weeks in Professor Michael Ferguson's laboratory in Dundee; the help he gave me and the interest he has taken in the CD59 project have been invaluable.

I would like to thank the Chemistry Department at Open University who renewed my status as a PhD student, and especially my supervisor, Dr. Peter Taylor. In particular I have valued the time he has taken to become involved in and enthusiastic about a field which is lateral to his own. The perceptive, fundamental questions he has asked have often made me think more about concepts I might otherwise have taken for granted. Most of all I would like to thank Professor Raymond Dwek for all the facilities that are available within the GBI, for his generosity in allowing me the privilege to work for and to write this thesis and for all the time he has given me over many years. Without his knowledge and experience, his skills as a teacher and supervisor, and his constructive criticism and continual challenges which have enabled to me to improve and develop my thinking, this work could not have been accomplished.

Pauline Rudd
The Glycobiology Institute, Oxford

Easter 1995

The Structure and Function of Glycoforms

Contents

Glycoform Structure and Function

Glycoproteins generally consist of populations of glycosylated variants of a single protein (glycoforms). While the potential oligosaccharide processing pathways available to a protein are dictated by the cell in which it is expressed, the final glycosylation pattern reflects constraints imposed by the three dimensional structure of the individual protein. This thesis explores the extent to which a protein controls its own glycosylation by comparing the glycosylation patterns of soluble forms of leucocyte surface antigens (CD2, CD5d1, CD59, CD5/CD4d3,4 and CD48) expressed in Chinese Hamster Ovary cells (ch.2). In chapter 3 the N- and O-linked and anchor glycans of human erythrocyte CD59 were characterised; the data suggested that the individual protein may influence the processing of glycosylphosphatidyl inositol anchor glycans. In chapter 4-7 consequences of variable glycosylation were explored by preparing defined sets of glycoforms for testing in functional assays. Individual glycoforms of RNase B reduced the susceptibility of the protein to proteases and modulated the activity of the enzyme (ch.4). IgG rheumatoid factor, from patients with rheumatoid arthritis, contained a specific subset of IgG glycoforms with Fc oligosaccharides terminating in GlcNAc (IgG0) and an extra-glycosylated heavy chain. Exposed GlcNAc residues on multiply presented IgG0 bound serum mannose binding protein, suggesting a role for Fc sugars in recognition (ch.6). In chapter 7 a subset of highly active tissue plasminogen activator (tPA) glycoforms, type II dimers, was identified. In the fibrin stimulated activation of plasminogen, glycosylation of tPA site 184 (type I glycoforms) decreased kcat, while the Km was decreased by glycosylation at the variably occupied plasminogen site, Asn289. Chapter 8, which deals with technology, explores applications of HPLC and mass spectrometry to the analysis of oligosaccharides and in chapter 5 capillary electrophoresis was used to resolve transferrin glycoforms, suggesting a rapid diagnostic for carbohydrate deficient glycoprotein syndrome.

Abbreviations

A list of relevant abbreviations is also included at the beginning of each chapter.

A-chain: residues 1-275 of tissue plasminogen activator;

2AB: 2-aminobenzamide;

AHM: anhydromannitol;

ANTS: 8-aminonapthalene 1,3,6-trisulphonate;

APAM: *Aspergillus phoenicus* α1-2 mannosidase;

B- Bisecting N-acetyl glucosamine;

B-chain residues 276-527 of tPA;

BM: Bowes Melanoma;

C: Constant;

CAT: Cation Exchange Chromatography;

CD: Cluster Differentiation;

CDGS: Carbohydrate Deficient Glycoprotein Syndrome;

CDR: Complementarity Determining Region;

CE: Capillary Electrophoresis;

CHO: Chinese Hamster Ovary;

CRD: Carbohydrate Recognition Domain;

C-type lectin: Calcium Dependent Lectin;

CZE: Capillary Zone Electrophoresis (free solution CE);

DC: Diffusion Coefficient

Endo D: Endoglycosidase D;

ER: Endoplasmic Reticulum;

Fab: Antigen Binding Fragment of IgG;

FACE: Fluorophore Assisted Carbohydrate Electrophoresis;

F, Fuc: Fucose;

Fc: crystallisable Fragment of IgG;

G2,G1,G0: complex biantennary oligosaccharides containing 2,1 or 0 arms respectively terminating in galactose;

GC: Gas Chromatography;

GCMS: Gas Chromatography/Mass Spectrometry;

GlcNAc: N-acetyl glucosamine;

GnT: GlcNAc transferase;

GPI : Glycosylphosphatidyl Inositol;

gu: glucose units;

H (in formulae): Hexose;

H: Heavy chain;

HCF: Human Colon fibroblast;

J: non-glycosylated transferrin;

JBAM: Jack bean α–mannosidase

L: Light chain;

IEF: Iso-Electric Focusing;

IgG: Immunoglobulin G;

IgSF: Immunoglobulin superfamily;

M: denotes sample was prepared in cell lines in Monsanto Co.;

M (in formulae): Mannose

MALDI MS: Matrix Assisted Laser Desorption/Ionisation Mass Spectrometry;

MAC: Membrane Attack Complex;

MBP: Mannose binding protein;

MASP: mannose binding protein - associated serine protease;

MC127: Murine cell line;

MRRF: Molar Relative Response Factor;

N (in formulae): N-acetyl hexosamine;

N: neutral glycans; NA1, NA2, NA3, NA4: glycans with 1,2,3 or 4 negative charges respectively;

NDV; Newcastle Disease Virus;

NMR: Nuclear Magnetic Resonance;

NHT: Normal Human Transferrin;

OGS: Oxford GlycoSystems;

PA: Plasminogen Activator;

PAI: Plasminogen Activator Inhibitor;

PNGase F: Peptide N-glycosidase F;

P4 GPC: P-4 Gel Permeation Chromatography;

R: denotes sample was prepared in cell lines in the Rega Institute;

RA: Rheumatoid Arthritis;

RAAM: Reagent Array Analysis Method;

RF: Rheumatoid Factor;

RI: Refractive Index;

RNase: Ribonuclease;

S0-S6: Subsets of glycoform populations of human transferrin which contain from 0-6 sialic acid residues attached to complex glycans;

ScR: Scavenger Receptor;

tPA: Tissue Plasminogen Activator;

UDP: Uridine Diphosphate;

uPA: Urokinase Plasminogen Activator;

V: Variable;

WAX: Weak Anion Exchange

Composition formulae of oligosaccharides use the following abbreviations: H: hexose, N; N-acetyl hexosaminidase; F: fucose; SA: sialic acid

Key to glycans:

 **N-Acetyl glucosamine (GlcNAc) [N]**

 **Fucose (Fuc) [F]**

 **Mannose (Man) [H]**

 **Sialic acid (SA)**

 **Galactose (Gal) [H]**

 **Glucose (G) [H]**

Chapter 1
Introduction to the structure and function of glycoforms

Publications associated with this chapter

1. Opdenakker, G., Rudd, P.M., Ponting, C.J. and Dwek, R.A. FASEB (1993) 7,14 1330-1337 Concepts and Principles of Glycobiology.

2. Rudd, P.M., Woods, R.J., Wormald, M.W., Opdenakker, G., Downing, A.K., Campbell, I. D. and Dwek, R.A. (1995) BBA - in press The effects of variable glycosylation on the functional activities of ribonuclease, plasminogen and tissue type plasminogen activator.

Abbreviations: ER: endoplasmic reticulum; UDP: uridine diphosphate; RNase: ribonuclease GPI: glycosylphosphatidylinositol, CDGS: Carbohydrate Deficient Glycoprotein Syndrome; RA: rheumatoid arthritis; NMR: nuclear magnetic resonance; tPA: tissue plasminogen activator; C-type lectin: calcium dependent lectin; CRD: carbohydrate recognition domain; CDR: complementarity determining region.


N-Acetyl glucosamine (GlcNAc) [N]


Fucose (Fuc) [F]


Mannose (Man) [H]


Sialic acid (SA)


Galactose (Gal) [H]


Glucose (G) [H]

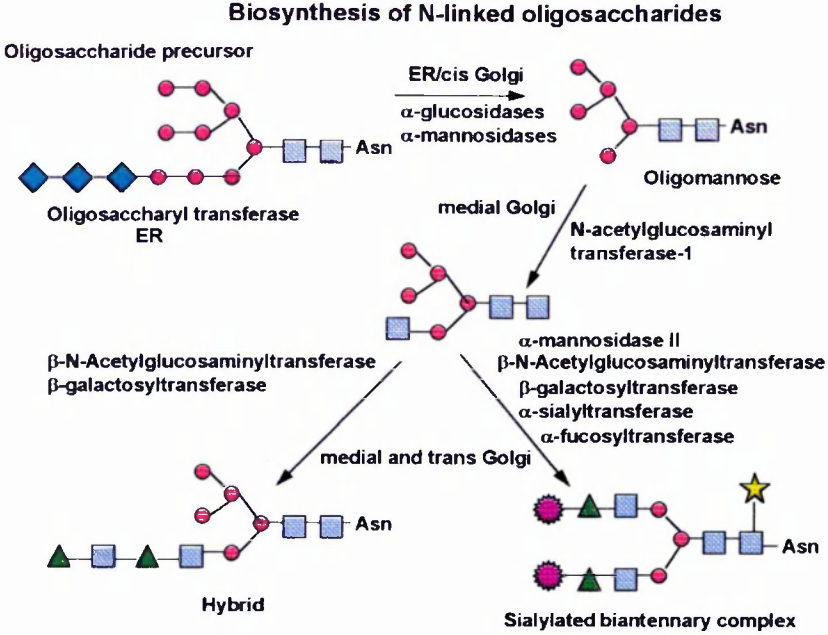


Figure 1: A summary of the glycosylation processing pathway.

Glycosylation is initiated in the rough ER where the membrane bound oligosaccharide transferase mediates the transfer of the oligosaccharide precursor, $Glc_3Man_9GlcNAc_2$, to the nitrogen in the side chain of some asparagine residues. Processing of the glycans begins in the ER by the enzymatic trimming of the three glucose residues and one of the mannose residues. The glycoproteins are then transported by transporter proteins in membranous vesicles to the Golgi apparatus for further processing. In the trans Golgi the completed glycoproteins are sorted for their final destinations such as the lysosomes, the plasma membrane or external secretion. Some of the enzymes involved and representative examples of the three classes of N-linked glycans and are shown: microheterogeneity within each structural class yields many more structures than are shown here. (Taken from Natuska, S. and Lowe, J.B. 1994). For key see previous page.

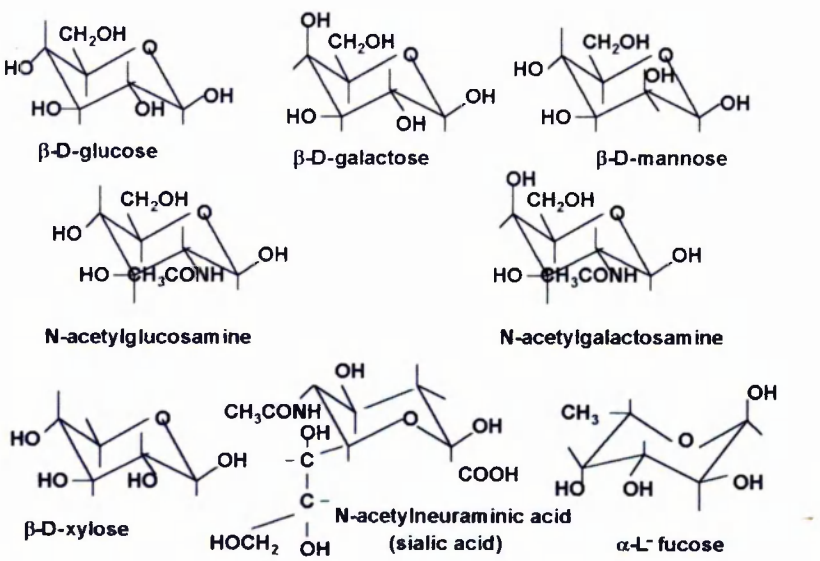


Figure 2: Monosaccharide residues commonly found in N- and O-linked oligosaccharides attached to proteins in mammalian systems.

Chapter 1
Introduction to the structure and function of glycoforms

1.1 The N-linked glycosylation biosynthetic pathway

The N-linked glycosylation pathway is initiated co-translationally for polypeptides which contain the potential glycosylation sequon AsnXaaSer, where Xaa is any other amino acid except proline. Oligosaccharyl transferases catalyse the covalent attachment of the oligosaccharide precursor, $Glc_3Man_9GlcNAc_2$, to some potential glycosylation sites on the protein. The glycoproteins are transported through the lumen of the endoplasmic reticulum (ER) into the cis, medial and trans Golgi stacks. Each of these compartments contains a different combination of highly specific glycosylating enzymes (Fig.1) which are anchored in the subcellular membranes with their active sites facing the lumen. An oligosaccharide chain is modified when a glycosidase or glycosyltransferase enzyme interacts with a specific glycoform of the developing glycoprotein, mediating the transfer or removal of a monosaccharide (Fig. 2) at the non-reducing end of the glycan chain. Three major classes of N-linked glycans have been identified: oligomannose, complex and hybrid (Fig.1). The efficient binding and processing of a glycoprotein by an enzyme requires the correct oligosaccharide to be presented in the context of a protein with a suitable local three dimensional structure. When processing involves the addition of a monosaccharide by a glycosyl transferase the event also involves a donor sugar-nucleotide (Fig. 3). The glycoproteins which emerge from the trans-Golgi are the end products of a series of enzyme reactions in which the immature glycoprotein generated by one reaction becomes the substrate for the next. The reactions may not necessarily go to completion, therefore glycoproteins generally consist of mixtures of glycosylated variants or glycoforms.

1.2 O-glycans

All Ser(Thr) O-linked glycosylation is initiated by the incorporation of a single N-acetylgalactosamine residue from UDP-GalNAc into the polypeptide chain, and elongation requires the sequential addition of monosaccharide residues. At least six core structures have been identified (Fig. 4; Schacter and Brockhausen 1992). In general no consensus sequence is required for O-glycosylation, however it has been noted that proline residues are often close to an O-glycosylated Ser(Thr) residue, and that regions of peptides which contain a high proportion of serine, threonine and proline residues are frequently O-glycosylated. Also tPA (ch.7) contains an O-linked fucose residue attached to Thr61 in the growth factor domain (Harris et al 1991). Threonine 61 is part of a consensus sequence for O-fucosylation, Gly Gly Ser/Thr, which has been noted in other epidermal growth factor regions in other molecules associated with the blood coagulation cascade such as human factor VII,IX,XII and bovine Factor VII (Nishimura et al 1992). In this thesis O-linked glycans containing type 1 cores were identified on human CD5 domain (d) 1 and a chimera of CD5 and CD4d3,4 expressed in CHO cells, and also on human erythrocyte CD59 and neutrophil gelatinase B.

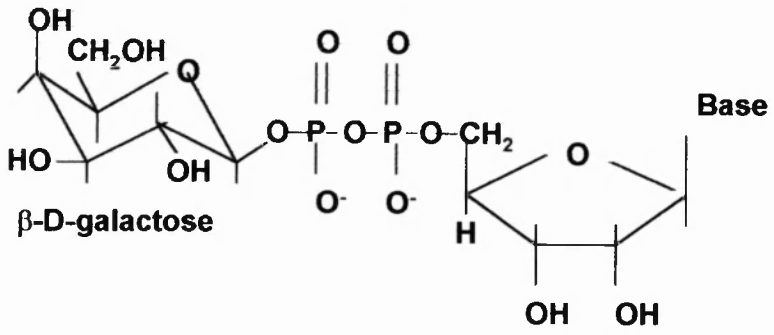


Figure 3: Galactose uridine diphosphate.

Each monosaccharide is linked to a sugar nucleotide from which it is transferred to the glycan chain by a glycosyl tranferase. The sugar nucleotides and some of their corresponding monosaccharides are as follows: UDP: GlcNAc, GalNAc, Gal, Glc, Xyl; GDP: Fuc, Man; CMP: N-acetyl neuraminic acid (sialic acid).

Core 1 :	Galβ1-3 GalNAc-R
Core 2 :	GlcNAcβ1-6 (Galβ1-3) GalNAc-R
Core 3 :	GlcNAcβ1-3 GalNAc-R
Core 4 :	GlcNAcβ1-6 GlcNAcβ1-3 GalNAc-R
Core 5 :	GalNAcα1-3 GalNAc-R
Core 6 :	GlcNAcβ1-6 GalNAc-R (R is α Ser/Thr)

Figure 4: The O-glycan core structures

1.3 Glycosylphosphatidyl inositol membrane anchors

Glycosylphosphatidyl inositol (GPI) anchors are complex glycolipids found covalently attached to a wide variety of cell surface plasma membrane proteins (Ferguson 1991, McConville and Ferguson 1993). A protein destined to receive a GPI anchor must contain an N-terminal signal sequence for entry into the lumen of the endoplasmic reticulum where the pre-assembled anchor precursor is attached. The C-terminus of the protein must contain a GPI signal sequence, the most common feature of which is a series of 12-20 hydrophobic residues at the C-terminus of the primary translation product preceded by a polar region of amino acids. This signal sequence is cleaved and replaced by a pre-assembled GPI precursor. The C-terminal cleavage amino acid (restricted to either Cys, Asp, Asn, Gly, Arg or Ser) is linked via ethanolamine phosphate to a glycan with a conserved backbone sequence (Manα1-2Manα1-6Manα1-4GlcNH$_2$) which may be further processed. The backbone is linked to the 6-position of the *myo*-inositol ring of phosphatidyl inositol (PI). GPI anchors contain two lipids, normally one acyl and one alkyl, attached through the phosphate on the inositol ring. In addition, a subset of anchors are palmitoylated at C3/4 on the inositol ring. In this thesis the glycosylation of the GPI anchor attached to human erythrocyte CD59 (Fig. 5) has been determined (ch.3).

1.4 Variable sequon occupancy

Although the same glycosylation machinery is available to all the proteins which are translated in a particular cell and use the secretory pathway, it has been estimated that between 10% and 30% of potential glycosylation sites are not occupied (Gavel and Heijne, 1990, Mononen and Karjalainen 1984). At the cellular level two factors which control the proportion of occupied glycosylation sites are the availability of the oligosaccharyl transferase and the dolichol pyrophosphate linked Man$_9$GlcNAc$_2$Glc$_3$ precursor oligosaccharide. The lipid component of the precursor is dolichol, a long chain isoprenol of 14-20 isoprene units (Fig. 6) which anchors the growing oligosaccharide in the lipid bi-layer of the endoplasmic reticulum (ER). Initially the sugar chain is processed in the cytoplasm (Fig. 7a) by sequential additions of mannose. When it contains a Man 5 glycan the molecule 'flips' so that the remainder of the processing of the Man 5 glycan to the Man$_9$GlcNAc$_2$Glc$_3$ precursor takes place within the lumen of the ER. The precursor is transferred to the asparagine residue in the glycosylation sequon of the nascent glycoprotein by an oligosaccharyl transferase which is situated 30-40A above the membrane of the ER and oriented roughly parallel to the membrane surface (Nillson et al 1993). If the normal levels of either the transferase or the precursor are significantly altered the proportion of occupied glycosylation sites may change. For example, in oviduct tissue the level of oligosaccharyl transferase activity may control the extent of protein glycosylation (Singh et al 1981) and the addition of dolichol phosphate to tissue slices has been shown to increase the percentage of ribonuclease (RNase) B in bovine pancreatic RNase from 12-90% (Carson et al 1981). Under occupancy of N-linked sites on many glycoproteins is characteristic of Carbohydrate Deficient Glycoprotein Syndrome (CDGS) (Jaeken at al 1984) and this may reflect abnormalities in the glycosylation pathway. Human transferrin is a clinical marker for CDGS

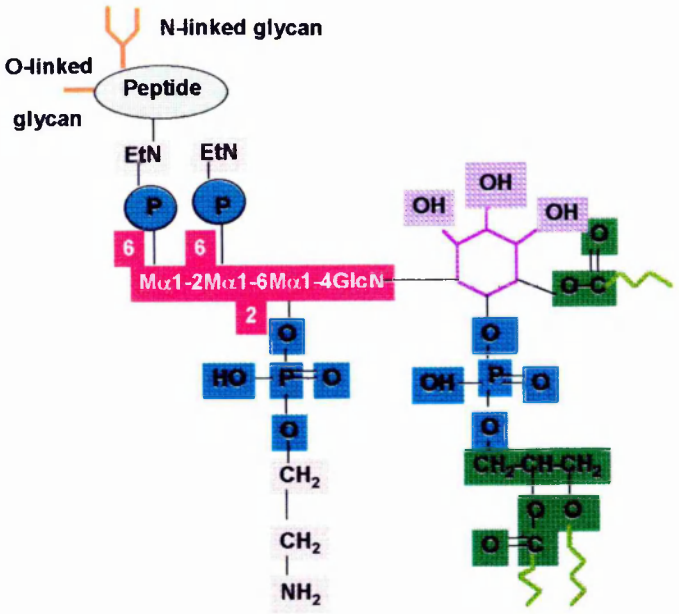
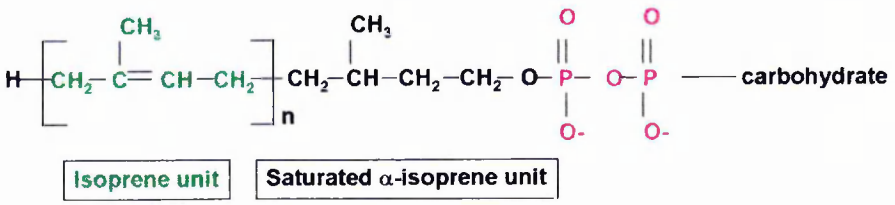


Figure 5: The human erythrocyte complement regulatory protein, CD59, contains a GPI anchor



Dolichol phosphate

Figure 6: The lipid component of the dolichol phosphate precursor oligosaccharide

and has two glycosylation sites, both of which are normally occupied. Some of these sites are not occupied in transferrin from CDGS patients, and the resulting abnormal glycoforms have been resolved and quantified in this thesis (ch.5). Site occupancy may also be controlled at the protein level. The three dimensional structure of the individual protein has a role in determining the extent of its glycosylation (Goochee et al 1992). A number of mechanisms may be involved, including the position of the glycosylation site in the protein. N-linked sites at the exposed turns of β-pleated sheets, which may be close to proline residues, are normally occupied while those close to the amino- or carboxy termini are generally less efficiently glycosylated. For example, the cytokine interleukin-1β, which has a potential glycosylation site at Asn 7 Cys Thr, is 50% glycosylated when expressed in *Saccharomyces cerevisiae*. Increasing the distance between the Asn 7 residue and the signal processing site to at least nine amino acids resulted in almost 100% glycosylation of the sequon (Livi et al 1991). Another influential factor may be the local environment of amino acids which surround the sequon. In particular, folding or packaging of the protein which involves disulphide bridges may hinder glycosylation as is the case with IL-6. It has been shown that eliminating the disulphide bond between Cys 45 and Cys 51 increases the efficiency with which the sequon at Asn 46 is glycosylated (Hasegawa et al 1992). The polypeptide sequence may determine the speed with which protein folding renders the sequon inaccessible, and this may result in a 'competition' between the rate of folding and the addition of the dolichol linked precursor (Savvidou et al 1984). In general, glycosylation at one site does not determine the sequon occupancy of another. For example, the rabies virus glycoprotein has a sequon at Asn 37 Leu Ser which is not normally glycosylated; the glycosylation efficiency at this site is not significantly increased when the normally glycosylated sequons at Asn 247GluThr and Asn 319LysThr are deleted (Shakin-Eshleman et al 1992). In this thesis three enzymes all of which contain one variably occupied N-glycosylation site have been studied. The possibility that variable site occupancy may modulate functional activity has been examined in the natural glycosylation variants of ribonuclease (RNase A and B), tissue plasminogen activator (tPA type I and II) (Fig. 8) and plasminogen (type 1 and 2).

1.5 Cell specific glycosylation

Glycosylation allows the cell a means of modifying many different proteins without recourse to the genome, and to respond flexibly to both transitory changes which occur during the cell cycle, and developmental changes in the organism. For cells in stasis a steady state is assumed to prevail in each of the subcellular compartments and the final populations of glycoforms are reproducible and not random. This may change, as for example when cell culture conditions cause both enzyme and substrate levels to vary. The cells respond to the changing conditions by modulating the expression of glycoproteins, glycosylating enzymes and sugar nucleotides. Different cell lines contain different repertoires of glycosylating enzymes and may process the glycans attached to the same protein in a highly specific manner. For example, Thy-1 expressed in rat brain has no glycoforms in common with the same protein

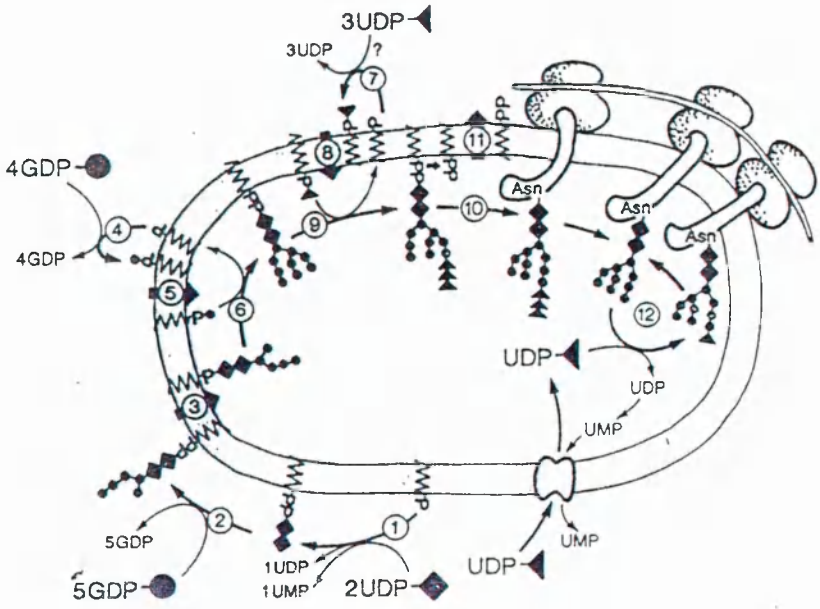


Figure 7: The synthesis of the dolichol phosphate precursor. *This takes place first on the cytoplasmic face and secondly on the lumen face of the ER.*

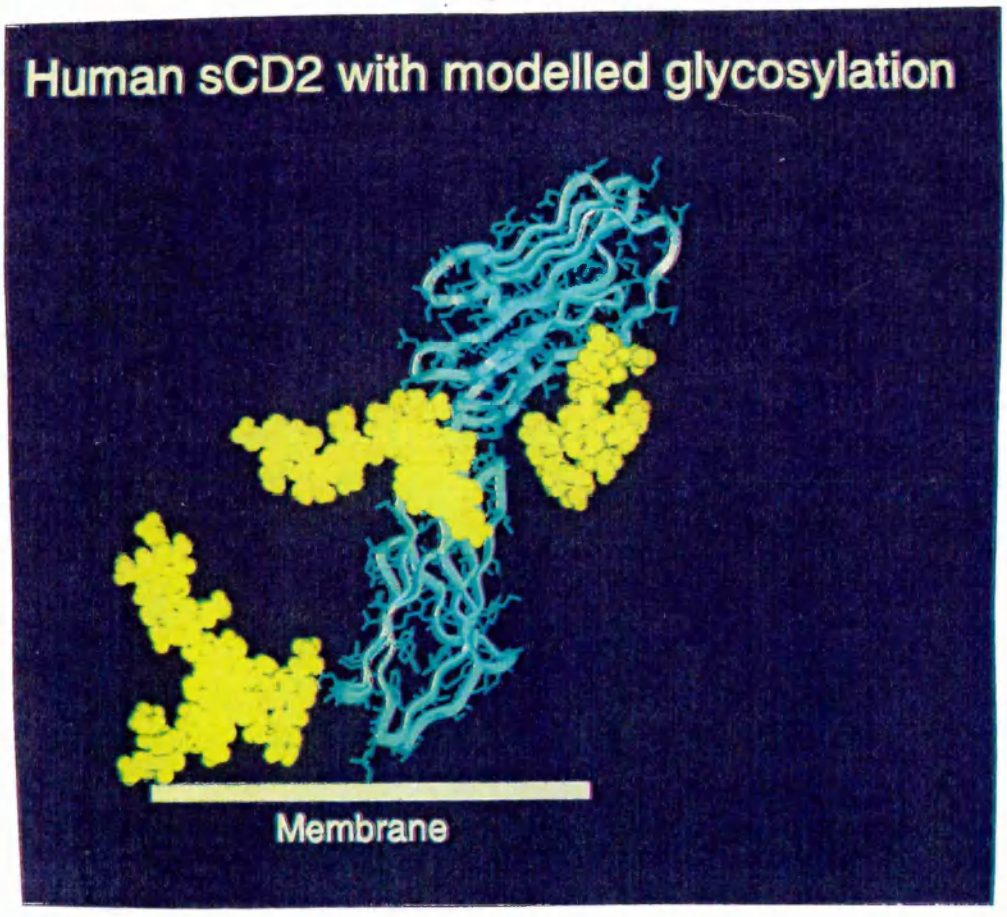


Figure 7b: Human CD2 shows site specific glycosylation

Asn 65 in domain 1 (upper domain) contains exclusively oligomannose sugars (Recney et al 1992) while the two remaining sites in domain 2 (lower domain) contain complex and oligomannose glycans (ch.2c).

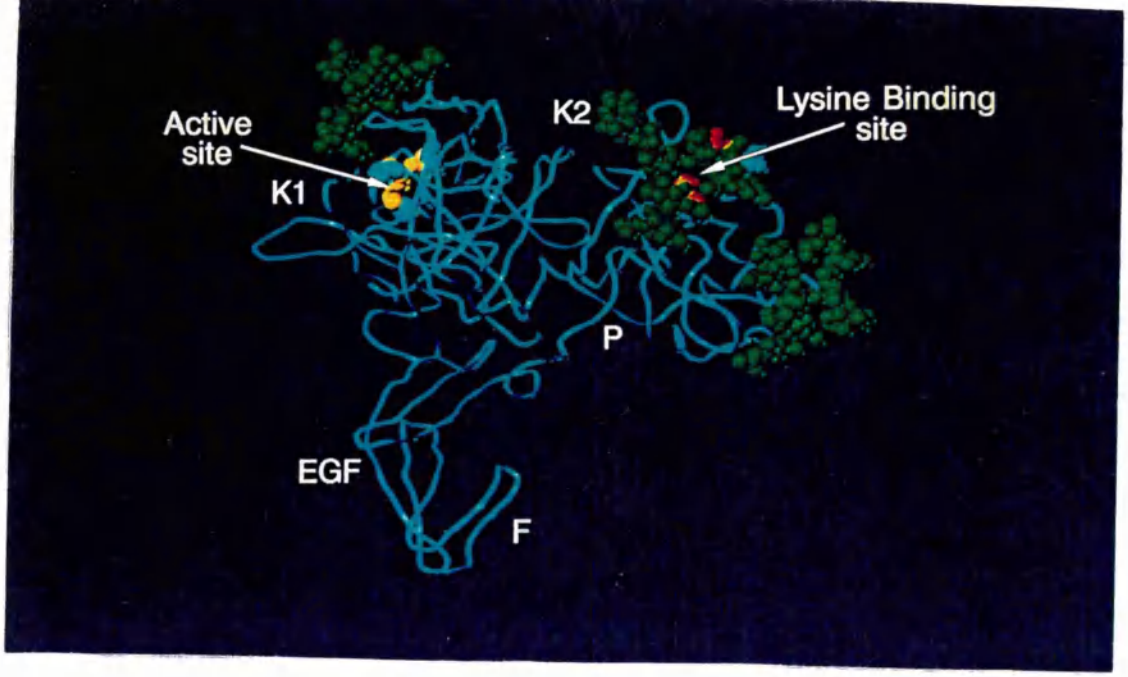
expressed in rat thymus (Parekh et al 1987). In this thesis differences in cell specific glycosylation machinery were explored by mapping the glycoform patterns of one protein, tissue plasminogen activator, expressed in ten different cell cultures using five different cell lines (ch.7). In mammalian systems polyclonal glycoproteins are secreted by mixtures of clones of cells which have many different sets and ratios of glycosyl transferases, and changes in the observed glycosylation patterns of such proteins may result from expansions in particular populations of cells. For example, the presence of increased levels of the agalactosyl glycoform population of polyclonal serum IgG in rheumatoid arthritis (RA) is well established (Parekh et al 1985) and has been noted in this thesis (ch.6). It has been proposed (Furukawa et al 1989) that agalactosyl IgG glycoforms in RA are secreted by a specific set of B-cell clones in which the IgG specific galactosyl transferase has decreased activity and affinity for UDP-galactose compared with that in normal B-cells.

1.6 Site specific glycosylation

A single multiply glycosylated glycoprotein starting in the endoplasmic reticulum with identical glycans may contain non-identical structures at each site after processing. For example, in rat brain Thy-1 site 23 (Asn-Asn-Thr) contains only oligomannose sugars while site 74 (Asn-Phe-Thr) contains both complex and hybrid structures and site 98 (Asn-Lys-Ser) is associated with complex, hybrid and oligomannose glycans (Parekh et al 1987). This indicates that, at each site, the combination of the local protein structure and the attached glycan constitutes a different structure as substrate for the processing enzymes. Another example of a glycoprotein which shows site specific glycosylation is CD2 (Fig7b). CD2 has three N-glycosylation sites, one of which (Asn 65), contains exclusively oligomannose structures (Recny et al 1992) while the remaining two sites (Asn 116 and Asn126) contain a mixture of oligomannose and complex glycans (ch.2). A possible reason for the conservation of glycosylation at Asn 65 has been explored in this thesis. IgG shows site specific glycosylation; for example Fab contains more sialic acid than Fc and in RA the Fc contains higher levels of glycans terminating in N-acetyl glucosamine than those attached to Fab (ch.6).

1.7 The generation of glycoforms

An oligosaccharide chain attached to a protein is modified when a glycosidase or glycosyl transferase enzyme interacts with a specific glycoform of a glycoprotein. The structure of the developing glycoprotein is an important factor which modulates the efficiency of each reaction, and the reaction rates for a single enzyme interacting with a range of different glycoproteins and glycoforms varies according to the affinity of the enzyme for the substrate. At any time many different glycoproteins may be in the glycosylating pathway, simultaneously competing for the active sites of the enzymes. Reaction rates also depend on the relative concentrations of enzyme to substrate. If these are not optimal some glycoprotein substrates in the pathway may be not processed, allowing variations to arise in the glycosylation of different proteins and microheterogeneity to develop in the glycoform population of a single glycoprotein. Another factor which increases the potential for heterogeneity is suggested by the finding that some



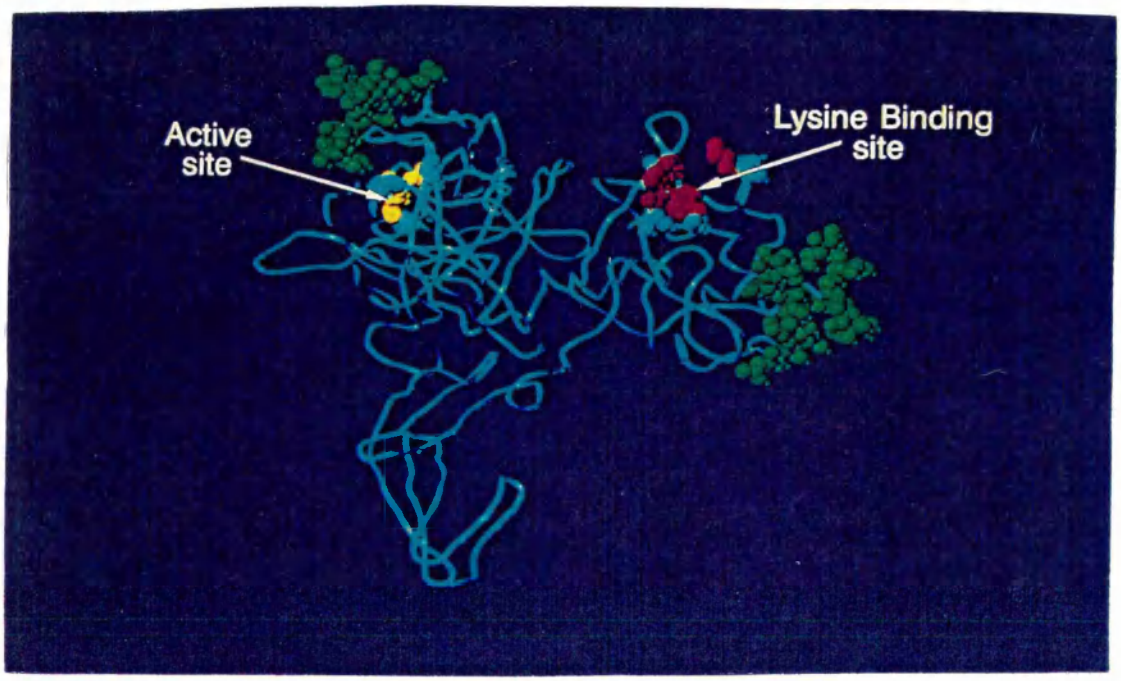


Figure 8: Schematic model of tissue plasminogen activator type I (above) and II (below).

Tissue plasminogen activator is composed of five domains of domains: the fibronectin type 1 'finger' module, the epidermal growth factor-type (EGF) domain, kringles 1 and 2 and the catalytic serine protease domain. This model was constructed using the co-ordinates of the finger and growth factor pair (Smith et al 1995), and kringle 2 (de Vos et al 1992) from human tPA. Kringle 1 has been modelled by homology. The figure illustrates the relative sizes of the sugars with respect to the protein and the site specific glycosylation. The oligomannose sugars at site Asn 117 and complex structures at Asn184 and Asn 448 are shown in green. The lysine binding site and the catalytic triad are highlighted in red and yellow respectively. Type II t-PA lacks glycosylation at Asn 184. Comparison of the glycosylation of types 1 and 2 tPA suggests that glycosylation at site 184 may sterically interfere with lysine binding. (Rudd et al 1995)

glycosyltransferases are localised throughout the Golgi stacks (Paulson and Colley 1989). One of these is β-galactosyl α2,6-sialyltransferase which, by competing with β-N-acetyl glucosaminyl transferase for terminal galactose residues, initiates a competition between chain elongation and chain termination. As a result of many factors glycoproteins with a range of glycans at different stages of processing may be transported into the next subcellular compartment to encounter a different set of enzymes. Since a glycoprotein which fails to be processed by one enzyme cannot become the substrate for the next, modification of the glycan chains will terminate whenever the developing glycoform is unable to interact further with the specific glycosylation machinery available in the cell. The glycoproteins which emerge from the trans-Golgi are the end products of a series of enzyme reactions in which the immature glycoprotein generated by one reaction becomes the substrate for the next. Glycoproteins, therefore, may be differently glycosylated from each other even when expressed in the same cell (ch.2) and moreover single glycoproteins generally consist of mixtures of glycosylated variants or glycoforms.

1.8 Key steps in the glycosylation pathway determine the class of glycan

Several key steps are involved in the biosynthesis of N-linked oligosaccharides; together these determine which of the potential glycosylation sites in the protein will be occupied, the class of the oligosaccharides (oligomannose, complex or hybrid) located at each site, and the terminal residues on the sugars. The efficiency of these key enzyme reactions and those which refine the basic structures in intermediate glycosylation processing, determine the final structures of the glycans. Initiation of glycosylation in the ER is mediated by an oligosaccharyl transferase which transfers the oligomannose precursor to the potential glycosylation site in the peptide. The pattern of cleavage of the mannose residues in the oligomannose precursor, $Glc_3Man_9GlcNAc_2$, by α-mannosidase I controls the proportions of Man 6-Man9 oligomannose glycoforms which will be in the final mixture of glycosylated variants. Processing of the oligomannose Man 5 glycans to hybrid structures in the Golgi follows the trimming of Man 8 oligomannose structures by α-mannosidase I to Man5. When presented in the context of an appropriate protein structure this glycan is a substrate for N-acetyl glucosaminyl transferase I which may substitute N-acetyl glucosamine on the α1,3 arm of the glycan. Branching to multiantennary structures requires the action of a series of N-acetyl glucosaminyl transferase enzymes (Schacter 1986), one for each antenna (Fig. 9). The number of antennae in mammals ranges from 2-4 but hen oviduct, which contains GlcNAc transferase VI, can make a penta-antennary structure. Processing to complex sugars depends on the removal of the two remaining terminal mannose residues on the α1,6 arm of the glycan by α-mannosidase II. The product of this reaction is a potential substrate for β-N-acetyl glucosaminyl transferase II which may add N-acetyl glucosamine to the 1,6 arm of the glycan, committing the glycoprotein to the complex sugar processing pathway. This pathway may terminate after the addition of galactose to one or both arms of the glycan or continue to extend the complex glycan by the addition of polylactosamine units (Galβ1,4GalNAc). Termination steps may be regulated by enzyme : substrate ratios, and oligomannose structures may be terminated at any stage of the trimming

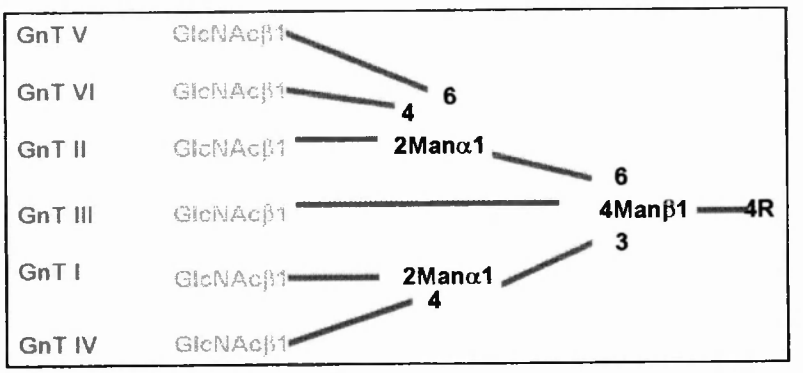


Figure 9: The GlcNAc transferases required for branching. *Five antennae can be initiated on the Manα1,6(Manα1,3) Manβ1-4GlcNAcβ1-4GlcNAcβ-Asn core of N-glycans by the GlcNAc transferases I, II, IV, and V. A bisecting GlcNAc can be added by GlcNAc tranferase III. (Figure taken from Oxford GlycoSystems catalogue 1994)*

process which reduces Man 9 structures to Man 5. Complex or hybrid structures contain processed arms which terminate in galactose, N-acetyl glucosamine or sialic acid.

1.9 Many enzymes are required to process the oligosaccharides on a glycoprotein

Each of the key steps requires the interaction of the glycoprotein with one or more highly specific enzymes (Reviews: Natsuka and Lowe 1994, Schacter 1991). For example, 11 different enzymes are required to attach and process one of the most common N-glycans in mammalian systems, the sialylated core fucosylated biantennary complex N-linked glycan (Fig.1) demonstrating how many aspects of cellular structure and metabolism are committed to glycoprotein synthesis. It has been calculated (Natsuka and Lowe 1994) that there must be in the order of a few hundred distinct glycosyltransferases, and the corresponding genes, to synthesise the many different oligosaccharides identified in mammalian species. Some glycosyltransferases are themselves glycosylated, and many are processed into soluble forms by the action of proteases which release them from the Golgi membrane. While each glycosyl tranferase with its corresponding sugar nucleotide can catalyse the synthesis of only a single glycosidic linkage, multiple distinct enzymes can catalyse the synthesis of identical sugar linkages, for example five α1,3 fucosyl tranferase genes have been cloned and characterised. To date some 30 glycosyltransferases have been characterised. Despite the finding that there is very little, if any, primary sequence homology between enzymatically distinct glycosyltransferase families, the enzymes which have been cloned suggest that there is a common topology. In general, the enzymes consist of a short (<25 residues) amino-terminal cytoplasmic domain, a single transmembrane segment and a larger carboxyl- terminal catalytic domain (generally more than 325 residues) in the lumen of the Golgi apparatus. Superimposed on the direct enzyme regulated mechanisms is the modulation of substrate availability controlled by the 3-dimensional structure of the protein, and it is the combination of these factors which directs the glycosylation of each protein.

2. The role of oligosaccharides in modifying structure and functional activity and in recognition

Some general principles relating to the role of oligosaccharides which are relevant to this thesis are discussed in the following three sections. First, oligosaccharides modify the local structure and overall dynamics of the protein to which they are attached. Second, oligosaccharides may modify the functional activity of a protein and third, oligosaccharides may be involved in molecular recognition events (review: Rademacher et al 1988).

2.1 Oligosaccharides modify the local structure and overall dynamics of proteins

X-ray crystallography has shown that each region of homology in the IgG molecule corresponds to a compact, independently folded unit, and that these are linked together by short sections of polypeptide chain. Each domain consists of two β-pleated sheets with anti-parallel strands connected by loop regions. This general pattern has been termed the immunoglobulin fold (Amzel and Poljak 1979). Crystallographic studies on immunoglobulin Fc fragments

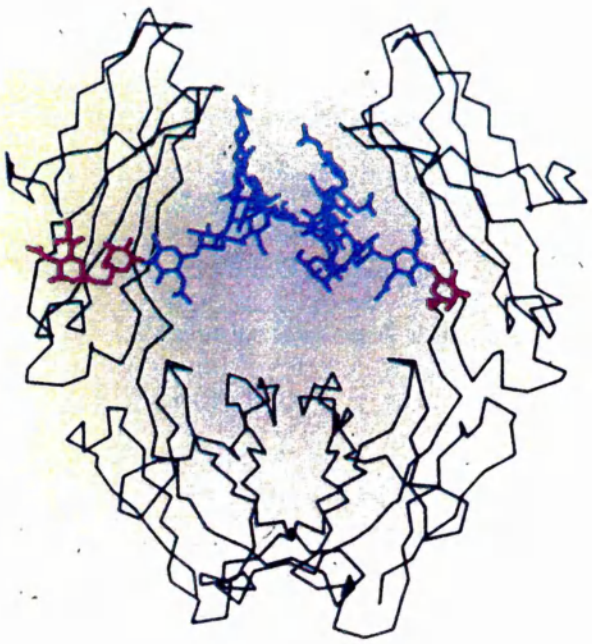


Figure 10: The Fc fragment of IgG

The model is based on the crystal structure of Diesenhofer (1981). The region between the two CH2 domains accommodates the oligosaccharides which are attached to Asn 297 on each heavy chain. The α1,3 arms of the glycans form a bridge across the domains while the α1,6 arms interact with the protein surface (ch6). The α1,6 arm of the complex biantennary oligosaccharide attached to the heavy chain on the left contains galactose (red) and terminates in sialic acid (red) while the glycan on the right terminates in galactose (red). In rheumatoid arthritis there are increased level of glycoforms in which both sialic acid and galactose are absent and which therefore terminate in N-acetyl glucosamine. In contrast to the C1q and Protein A binding sites, which are at the carboxy terminal side of Asn 297 (Phe 319 and Ile 332), the monocyte binding site involving residues 234-237 on the lower hinge region is markedly affected by the removal of carbohydrate.

(Diesenhofer 1981; Sutton et al 1983) have shown that, unlike other immunoglobulin domains, the two CH2 domains do not form extensive lateral associations. The resulting interstitial region accommodates the oligosaccharides, which are attached to Asn 297 on each heavy chain, such that the 1,3 arms of the complex carbohydrates form a bridge across the domains and the 1,6 arms interact with the protein surface (Fig. 10). The effects of glycosylation may be highly specific. Both the C1q and protein A binding sites on IgG are located in the CH2 domain between Phe 319 and Ile 332 (distal or at the carboxy terminal side from the N-linked glycosylation site). Neither of these functions is markedly affected by the absence of carbohydrate (Leatherbarrow and Dwek 1983; Leatherbarrow et al 1985). However the binding of monocytes, which involves residues 234-237 (proximal to the amino terminus) on the lower hinge region, is eliminated in aglycosylated monoclonal murine IgG. This may be because the absence of sugars results in a lateral movement of domains in the hinge region relative to the normally glycosylated antibody. Moreover, in aglycosylated IgG a protein structural change has been detected by 1H NMR at His 268, which is also in the vicinity of the lower hinge. These data suggest that the oligosaccharides may stabilise a particular hinge conformation essential for monocyte binding. The spatial relationship between the CH2 and CH3 domains, on the other hand, does not seem to depend on the presence of the oligosaccharides since, in their absence, protein A and C1q binding are unaffected.

Bovine pancreatic RNase contains one N-linked glycosylation sequon (Asn 34 Lys Thr). It occurs naturally as both the non-glycosylated form, (RNase A) and a mixture of glycosylated variants collectively known as RNase B. NMR structural studies (Joao et al 1992) have compared the hydrogen-deuterium solvent exchange rates for the NH protons of RNase A and B. The presence of the sugar decreases the exchange rate in 30 of the 124 amino acid residues suggesting that the overall dynamic stability of the molecule is enhanced such that it becomes more rigid. Some of the affected residues are close to the oligosaccharide attachment site, others are as far away as 30A, while three (Gly11, Asn 44, and His 12) are involved in the active site, which catalyses the cleavage of the phosphodiester linkages of RNA. The sugars attached to proteins are large compared to the domains to which they are attached. For example the glycans on tPA (Fig. 8) are approximately the same size as a kringle region (3nm), and a single monosaccharide unit (0.6nm) is approximately the same size as a tryptophan ring. The dynamics of the sugar and the flexibility of the Asn side chain allow oligosaccharides to shield large areas of the protein surface (Fig. 11, Woods et al 1994). Thus sugars may modify the dynamics, stability and functional activities of the proteins to which they are attached.

2.2 Oligosaccharides may modify the functional activity of a protein

Oligosaccharides may modify the functional activity of an enzyme by steric hindrance of a functionally important site. Tissue plasminogen activator (t-PA) (Fig.8) is a serine protease glycoprotein consisting of five domains: a fibronectin 'finger' domain, an epidermal-growth factor domain, two kringles and the catalytic serine protease domain (Ny et al 1984, Patthy 1985). Binding sites for fibrin are present within the 'finger' and kringle domains. There are two main classes of glycosylated variants of the molecule, Type I and Type II; Type I has three N-

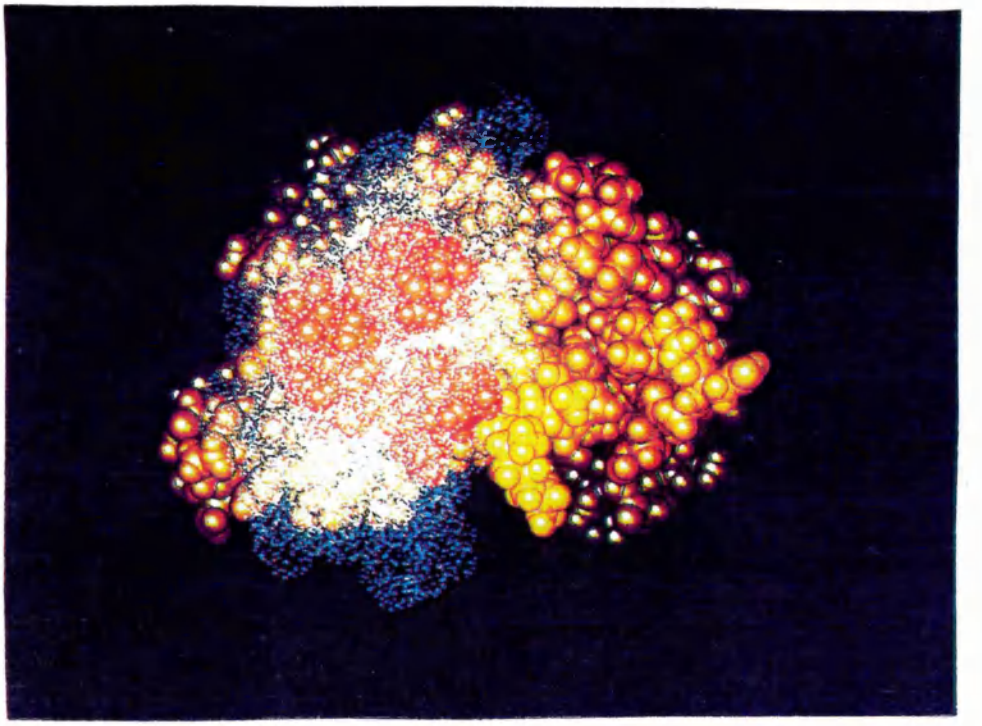


Figure 11: The Man 9 oligomannose oligosaccharides attached to RNase B

The dynamics of the sugar and the flexibility of the Asn side chain allow the oligosaccharides to shield large areas of the protein surface (Woods et al 1994). The model shows the effect of flexibility of the Asn-34 side chain on the orientation of the oligosaccharides attached to RNase B. The Man9 glycoform is shown based on the crystal structure (yellow CPK spheres) (Williams et al 1987) with an overlay of 10 oligosaccharide conformations. The total van der Waals surfaces of the oligosaccharides are illustrated by white dots.

linked sugars, at Asn 117, Asn 184 and Asn 448, while Type II has only two, at Asn 117 and Asn 448. In addition to structural differences in t-PAs derived from various cell lines, functional differences have been established by comparing the activity of Type I and Type II in assays which involve the kringle 2 domain in which the variably occupied sequon at site 184 is situated (Parekh et al 1989a, Parekh et al 1989b, Wittwer et al 1989). A two - fold reduction in the rate of t-PA mediated fibrin-dependent plasminogen activation has been observed. This may be the result of steric hindrance by the sugar of the kringle 2 lysine binding site which binds exposed lysine sites on partially degraded fibrin.

2.3 Oligosaccharides may be involved in molecular recognition events

Oligosaccharides may confer an orthogonal property on the protein, such as a mechanism for clearance, without affecting the other functions of the molecule. There are a number of clearance mechanisms, some of which are sugar mediated. For example, the hepatic asialo receptor recognises glycoproteins with terminal galactose residues which have become exposed as a result of de-sialylation by neuraminidase. Such a general mechanism allows the cell to attach the same oligosaccharide recognition units to many different proteins without the need to 'code' this in the DNA for the protein. Many of the biological functions of carbohydrates are mediated through recognition by a receptor. One of the most important classes of receptor is the calcium-dependent animal lectins (C-type lectins) (Drickamer 1988; Hoppe and Reid 1994; Weiss and Drickamer 1994). The prototype of this family is the asialoglycoprotein receptor, found on mammalian hepatocytes, which mediates clearance of glycoproteins. The receptor protein contains a single carboxy-terminal carbohydrate recognition domain (CRD). C-type CRDs have been found in a large number of proteins in association with a variety of effector domains. They are approximately 120 amino acids long and are characterised by 31 invariant or highly conserved amino acids. C-type lectins may therefore mediate carbohydrate-recognition in many different processes. X-ray data on the CRD from the mannose binding protein illustrates the type of recognition involved in the binding of oligosaccharides. The CRD shows specificity for terminal residues including GlcNAc, mannose, glucose and fucose and the interaction involves chelation of the sugar to the Ca^{2+} ion via cis hydroxyl groups (Fig. 12). There are also interactions between the CRD and the monosaccharide residues in the oligosaccharide which are mediated by water molecules - effectively increasing the surface area of the oligosaccharide in contact with the CRD. C-type lectins and their constituent CRDs display weak affinity for monosaccharides. The specificity for multi-branched oligosaccharides results from clustering of several CRDs, usually by the formation of oligomers of polypeptide chains each of which contains a single CRD.

Another important system which can involve oligosaccharide recognition is defence. Both the antibody molecule and the mannose binding protein recognise and bind oligosaccharides attached to pathogens, a step which may lead to their elimination. Some details of the antibody recognition of carbohydrate antigens have been worked out in a study of the binding of the antibody molecule to a cell surface oligosaccharide of pathogenic *Salmonella typhimurium* of

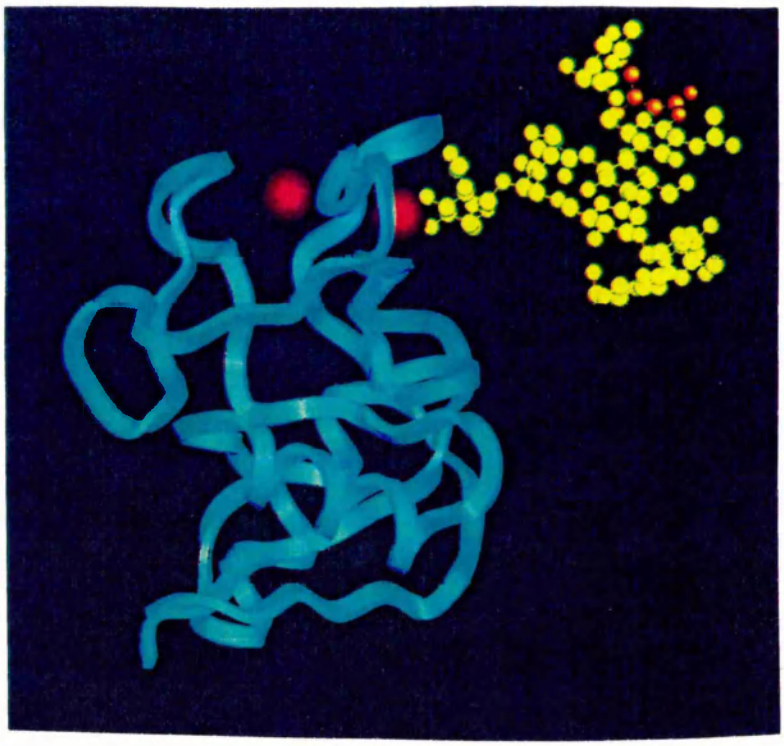


Figure 12: The mannose binding protein

A single carboxy-terminal carbohydrate recognition domain of the calcium dependent lectin mannose binding protein binding an oligosaccharide terminating in N-acetyl glucosamine (Wormald, M. - unpublished data based on Weiss et al 1992). The sugar chelates to the Ca^{2+} ion (red) via the 3- and 4- hydroxyl groups. The structure is stabilised by interactions between the CRD and the monosaccharide residues in the oligosaccharide which are mediated by water molecules.

serogroup B involves all these types of interactions (Cygler et al 1991). These indicate that the antibody binding site contains specific sub-sites for individual monosaccharide units. Binding to these sub-sites involves four distinct types of interaction all of which may contribute to the thermodynamics of association of a neutral oligosaccharide with an antibody combining site. First, the precision with which the oligosaccharide fits the binding site determines the residues in the complementarity determining region (CDR) that will contribute significant van der Waals forces to the binding. Second, amino acids with polar side chains, such as His and Glu, may interact with hydroxyl groups on the sugars. Third, monosaccharides in particular conformations can present a planar hydrophobic face which can interact with the planar rings of aromatic amino acids such as Phe, Tyr and Trp. Often these are in a position to be stacked against aromatic regions of monosaccharides, in a sandwich-like arrangement. Finally, water molecules, sometimes bound to amino acid residues buried deep in the structure, can form hydrogen bonds linking protein to sugar. The sugar hydroxyls themselves can stabilise particular conformations of the oligosaccharide by forming hydrogen bonds between adjacent monosaccharide units. Oligosaccharide binding to the antibody molecule is therefore a result of a combination of events which allow individual monosaccharide subunits to bind in specific sites on the protein. Although the affinity of each interaction may be low, high affinity as well as high specificity are achieved from the presence and relative orientation of several subsites and their relative presentations.

In contrast, serum mannose-binding protein is a C-type lectin which mediates antibody-independent recognition of pathogens which contain a high concentration of mannose or N-acetylglucosamine on their surface. This recognition is followed either by complement fixation or by opsonization. The clustering of CRDs together in the mannose binding protein allows recognition in high affinity of multi branched oligosaccharide ligands such as occur in yeast mannans which contain a main α-(1,6) linked chain with short branches of 1-3 mannose units linked either α–(1,2) or α-(1,3).

Oligosaccharides are frequently located on cell surface proteins where they are exposed and available to act as recognition markers. Some recent studies of lectin-carbohydrate interactions have suggested a number of reasons to suggest that oligosaccharides may be well adapted to mediate cell-cell or protein-protein interactions. The crystal structure of wheat germ agglutinin-glycophorin sialoglycopeptide receptor complex (Wright et al 1992) indicates that a single multivalent tetrasaccharide can bind several protein ligands. In addition, the tetrasaccharide was found to adopt different conformations in different lattice environments demonstrating that oligosaccharides can respond to accommodate the spatial requirements of lectin-protein ligands. Oligosaccharide flexibility is also important in the binding of the lectin LOL-1, isolated from the seeds of *L.ochrus* (Bourne et al 1992). The sugar adopts an extended S-shape conformation when bound to the lectin. Two residues (Man4 and GlcNAc5) interact on either side of a phenyl ring, gripping it as a clamp while two more glycan residues fit into a partly hydrophobic cleft and participate in numerous interactions with the lectin. The X-ray data also shows that the lectin adapts to the sugar; the protein backbone was observed to move towards mannose in the formation of the complex, which was stabilised by hydrogen bonds. It has also

been proposed from a study of the presentation of the N-linked complex sugar on human leucocyte elastase (Walter et al 1992), that one role of core fucosylation may be to order the presentation of the oligosaccharide by interacting with the protein and restricting the movement of the oligosaccharide in the region of the protein - carbohydrate linkage.

3. Introduction to the chapters in this thesis

The glycans associated with a range of leucocyte cell surface antigens (CD5 domain 1, CD2, soluble CD59, CD48 and a chimera of CD5 domain 1 and CD4 domains 3,4) have been analysed (ch.2). There were several reasons for this study. In the first place, the glycoproteins were all expressed in CHO cells and this allowed a study to be made of the extent to which the individual proteins direct their own glycosylation within a common biosynthetic pathway. Second, the glycans from CD2 and soluble CD59 were analysed to provide additional information for the X-ray and NMR studies which have been carried out on this material. Third, this sequencing project provided an opportunity to explore some new technology and develop some alternative strategies to those used in conventional approaches to oligosaccharide analysis (ch.8). The glycosylation of CD59 from human erythrocytes was also examined and the glycans within the GPI anchor compared with those N-linked to the protein (ch.3). Examination of a molecular model explored a role for the N-linked sugar in orienting the active site towards the binding site in C8 and or C9 which has been proposed by Ninomiya et al (1992). In general, many functions for sugars have been established both in recognition and in modifying functional activity, however the consequences of the arrays of glycoforms which are associated with individual glycoproteins have seldom been addressed. In the second part of this thesis the technical difficulties (ch. 8) involved in isolating and analysing single glycoform populations have been addressed in a number of systems (Ch.4-7). Individual sets of glycoforms were obtained either by modification of the natural population (RNaseB, transferrin, IgG) or by isolating natural glycosylation variants (RNase A and B, tPA types I and II, and rheumatoid factor IgG). Individual pure glycoforms or specific subsets of glycoforms were analysed, isolated and examined in a variety of assays. Single, pure glycoforms of RNase were prepared and examined for their ability to confer stability on the protein and to shield areas of the protein surface (ch.4). The glycoforms of transferrin were modified using exoglycosidase enzymes and variations in the glycoform populations with disease were monitored by capillary electrophoresis (ch.5). Pure glycoforms of IgG, Fab and Fc were prepared to examine a role for specific sugars in recognition (ch.6). The activity of cell type specific sets of glycoforms of tPA was tested, revealing a subset with increased fibrin dependent and independent activity. Finally, the role of variable sequon occupancy in the modulating the kinetics of the interaction of tPA with plasminogen types 1 and 2 was explored (ch.7). The results suggested a possible role for glycosylation in modulating the protease cascade associated with remodelling of the extracellular matrix. As a preliminary step to probing this possibility the glycosylation of gelatinase B, an enzyme involved in this cascade, was determined (ch.8).

Figure 13: *The largest natural glycoform of bovine pancreatic ribonuclease B, RNase Man9 (Woods et al 1994). The natural glycoform population, containing a mixture of glycoforms with from 5-9 mannose residues attached through the chitobiose core to Asn34, was enzymatically modified to give the single, electrophoretically pure glycoforms of RNase Man5, Man1 and Man0 (ch.4).*

4. References to chapter 1:

Amzel, L.M. and Poljak, R.J. (1979) Ann. Rev. Biochem. 48 961-997 Three dimensional structure of immunoglobulins.

Baron, M., Norman, D., Willis, A. and Campbell, I.D. (1990) Nature 345, 642-646 Structure and function of the fibronectin type 1 module.

Bourne, Y., Rouge, P. and Cambillau, C. (1992) J. Biol. Chem. 267 197-203 X-ray structure of a biantennary octasaccharide lectin complex refined at 2.3A resolution.

Carson, D.D., Earles, B.J., Lennarz, W.J. (1981) J. Biol. Chem. 256 11552-11557 Enhancement of protein glycosylation in tissue slices by dolichol phosphate.

Cooke, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Campbell, I.D., Gregory, H. and Sheard, B. (1987) Nature 327, 339-341 The solution structure of epidermal growth factor.

Cygler, M., Rose, D.R., and Bundle, D.R. (1991) Science 253, 442- 445 Recognition of a cell-surface oligosaccharide of pathogenic Salmonella by an antibody Fab Fragment.

Deisenhofer, J. (1981) Biochemistry 20, 2361-2370 Crystallographic refinement and atomic models of human Fc fragment and its complex with fragment B of protein A from *Staphylococcus aureas* at 2.9 and 2.8 -A resolution.

Drickamer K. (1988) J. Biol. Chem. 263, 9557-9560 Two distinct classes of carbohydrate recognition domains in animal lectins.

Ferguson, M.A.J. (1991) Current Opinion in Structural Biology 1 522-529 Lipid anchors on membrane proteins.

Furukawa, K., Matsuta, K., Takeuchi, F., Kosuge, E., Miyamoto, T. and Kobata, A. (1989) International Immunology 2 105-112 Kinetic study of a galactosyltransferase in the B-cells of patients with RA.

Gavel, Y. and von Heijne, G. (1990) Protein Engineering 3, 433-442 Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering.

Goochee, C.F., Gramer, M.J., Anderson, D.C., Bahr, J.B., and Rasmussen, J.R. (1992) The oligosaccharides of glycoproteins: factors affecting their synthesis and their influence on glycoprotein properties. In Frontiers in Bioprocessing II (Todd, P., Sikdar, K., and Bier, M. eds) pp. 199-240, American Chemical Society, Washington, D.C.

Harris, R.J., Leonard, C.K., Guzzetta, A.W. and Spellman, M.W. (1991) Biochemistry 30 2311-2314 Tissue Plasminogen Activator Has an O-linked Fucose Attached to Threonine-61 in the Epidermal Growth Factor Domain.

Hasegawa, M., Orita,T., Kojima,T., Tomonoh, K., Hirata,Y., and Ochi, N. (1992) Eur. J. Biochem. 210, 9-12 Improvement in the heterogeneous N-termini and the defective N-glycosylation of human interleukin-6 by genetic engineering.

Hoppe, H-J and Reid, K. (1994) Current Biology (Structure) 2 1129-1133 Trimeric C-type lectin domains in host defence.

Kitagawa H. and Paulson J.C. (1994) J. Biol. Chem. 269 1394-1399 Cloning of a novel α2,3 sialyl tranferase that sialylates glycoprotein and glycolipid carbohydrate groups.

Jaeken, J., van Eijk,H.G., van der Heul, C., Corbeel, L., Eeckels, R. and Eggermont E (1984) Clin. Chim. Acta. 144 245-247 Sialic acid deficient serum and cerebospinal fluid transferrin in a newly recognised genetic syndrome.

Joao, H.C., Scragg, I.G., and Dwek, R.A. (1992) FEBS Lett. 307, 343-346 Effects of glycosylation on protein conformation and amide exchange rates in ribonuclease B.

Leatherbarrow, R.J., Rademacher, T.W., Dwek, R.A., Woof, M.J., Clark, A., Burton, D.R., Richardson, N., and Feinstein, A. (1985) Mol. Immunol. 22 407-415 Effector functions of a monoclonal aglycosylated mouse IgG2a: binding and activation of complement component C1 and interaction with human monocyte Fc receptor.

Leatherbarrow, R.J. and Dwek, R.A. (1984) Molecular Immunology 21 321-327 Binding of complement subcomponent C1q to mouse IgG1, IgG2a and IgG2b: A novel C1q binding assay

Livi, G,P., Lillquist, J.S., Miles, L.M., Ferrara, A., Sathe, G.M., Simon, P.L., Meyers, C.A., Gorman, J.A., Young, P.R. (1991) J. Biol. Chem. 266, 15348-15355 Secretion of N-glycosylated interleukin-1b in *Saccharomyces cerevisia* using a leader peptide from Candida albicans. Effect of N- glycosylation on biological activity.

McConville, M.J. and Ferguson, M.A.J. (1993) Biochem. J. 294, 305-324. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes.

Mononen, I. and Karjalainen, E. (1987) Biochim. Biophys. Acta 788, 364-367 Structural comparison of protein sequences around potential N-glycosylation sites.

Nilsson, I.M. and von Heijne, G. (1993) J. Biol. Chem. 268 5798-5801 Determinations of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane.

Natsuka, S. and Lowe, J.B. (1994) Current Opinion in Structural Biology 4 683-691 Enzymes involved in mammalian biosynthesis.

Nishimura H, Takao T, Hase S, Shimonishi Y, and Iwananga S. (1992) J. Biol. Chem. 265, 17520-17525 Human Factor IX has a tetrasaccharide "O"-Glycosidically linked to Serine 61 through the Fucose Residue.

Ny, T., Eigh, F. and Lund, B. (1984) Proc. Nat. Acad. Sci. USA 81, 5355-5359 The structure of the human tissue-type plasminogen activator gene: correlation of intron and exon structures to functional and structural domains.

Parekh, R.B., Tse, A.G.D., Dwek, R.A., Williams, A.F., and Rademacher, T.W. (1987) EMBO J. 6, 1233-1244 Tissue-specific N-glycosylation, site specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1.

Parekh, R.B., Dwek, R.A., Sutton, B.J., Fernandes, D.L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, K., Takeuchi, F., Nagano, Y., Miyamoto, T. and Kobata, A. (1985) Nature 316, 452-457 Association of Rheumatoid Arthritis and Primary Osteoarthritis with changes in the glycosylation pattern of total serum IgG.

Parekh, R.B., Dwek, R.A., Thomas, J.R., Opdenakker, G., Rademacher, T.W., Wittwer, A.J., Howard, S.C., Nelson, R., Siegel, N.R., Jennings, M.G, Harakas, N.K., Feder, J. (1989) Biochemistry 28, 7644-7662 Cell-type-specific and site-specific N-glycosylation of type I and type II human tissue plasminogen activator.

Parekh, R.B., Dwek, R.A., Rudd, P.M., Thomas J.R., Rademacher, T.W., Warren T., Wun, T.-C., Hebert, B., Reitz, B., Palmier, M., Ramabhadran, and Tiemeier, D.C. (1989) Biochemistry 28, 7670-7679 N-glycosylation and *in vivo* enzymatic activity of human recombinant tissue plasminogen activator expressed in chinese hamster ovary cells and a murine cell line.

Patthy, L. (1985) Cell 41, 657-663 Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules.

Paulson, J.C., and Colley, K.J. (1989) J. Biol. Chem. 264 17615 Glycosyl transferases. Structure, location and control of cell specific glycosylation.

Rademacher, T.W., Parekh, R.B., and Dwek, R.A. (1988) Ann. Rev. Biochem. 57, 785-838 Glycobiology.

Recny, M.A., Luther, M.A., Knoppers, M.H., Neinhardt, E.A., Khandekar, S.S., Concino, M.F., Schimke, P.A., Francis, M.A., Moebius, U., Reinhold, B., Reinhold, V.N. and Reinherz, E.L. J. Biol. Chem. 267 22428-22434 N-glycosylation is required for human CD2 adhesion.

Rudd, P.M., Woods, R.J., Wormald, M.W., Opdenakker, G., Downing, A.K., Campbell, I. D. and Dwek, R.A. (1995) BBA 000 The Effects of Variable Glycosylation on the Functional Activities of Ribonuclease, plasminogen and tissue type plasminogen activator.

Savvidou, G., Klein, M., Grey, A.A., Dorrington, K.J., Carver, J.P. (1984) Biochemistry 23, 3736-3740 Possible role for peptide-oligosaccharide interactions in differential oligosaccharide processing at asparagine-107 of the light chain and asparagine-297 of the heavy chain in a monoclonal IgG.

Schacter, H. (1986) Biochem. Cell Biol. 64 163-181 Biosynthetic controls that determine the branching and microheterogeneity of protein bound oligosaccharides.

Schacter, H. (1991) Current Opinion in Structural Biology 1 755-765 Enzymes associated with glycosylation.

Schacter, H. and Brockhausen, I. (1992) In Allen, H.J. and Kisailus, E.C. (eds), Glycoconjugates. Composition, structure and Function. Marcel Dekker, Inc., New York, N.Y., 263-332

Shakin-Eshleman, S.H., Remaley, A.T., Eshleman, J.R., Wunner, W.H., and Spitalnik, S.L. (1992) J. Biol. Chem. 267, 15, 10690-10698 N-linked glycosylation of rabies virus glycoprotein.

Singh, B.N., Lucas, J.J. (1981) J. Biol. Chem. 256 12018-12022 Increased transfer of oligosaccharide from oligosaccharide pyrophosphoryl dolichol to protein acceptors upon estrogen-induced chick oviduct differentation.

Sutton, B.J. and Phillips, D.C. (1983) Biochem. Soc. Trans. 11, 130-132 The three-dimensional structure of the carbohydrate within the Fc fragment of immunoglobulin G.

de Vos, A.M., Ultsh, M.H., Kelley, R.H., Padmanabhan, K., Tulinsky, A., Westbrook, M.L. and Kossiakoff, A.A., (1992) Biochemistry 31, 270-279 Crystal structure of the kringle 2 domain of tissue type plasminogen activator at 2.4 A resolution.

Walter, J., Steigmann, W., Singh, T.P., Bartunik, H., Bode, W. and Huber, R. (1982) Acta Cryst B, 38, 1462 On the disordered activation domain in trypsinogen: chemical labelling and low temperature crystallography.

Weis, W.I., Drickamer, K., and Hendrickson, W.A. (1992) Nature 360,127-134 Structure of a C-type mannose-binding protein complexed with an oligosaccharide.

Weiss, W.I. and Drickamer, K. (1994) Current Biology (Structure) 2 1227-1240 Trimeric structure of a C-type mannose-binding protein

Wittwer, A.J., Howard, S.C., Carr, L.S., Harakas, N.K., Feder, J., Parekh, R.B., Rudd, P.M., Dwek, R.A., and Rademacher, T.W. (1989) Biochemistry 28, 7662-7669 Effects of N-glycosylation on *in vitro* activity of Bowes melanoma and human colon fibroblast derived tissue plasminogen activator.

Woods, R.J., Edge, C.J., and Dwek, R.A. (1994) Nature Structural Biology 1 499-501 Protein surface oligosaccharides and protein function.

Wright, C.S. (1992) J. Biol. Chem. 267 14345-14352 Crystal structure of wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex: structural basis for co-operative lectin binding.

Chapter 2

Protein specific glycosylation patterns of five leucocyte cell surface antigens expressed in Chinese Hamster Ovary cells.

Introduction

(A) The glycosylation of soluble human CD59

(B) The glycosylation of human CD5 d1 and a chimera of human CD5/rat CD4 d3&4

(C) The glycosylation of human CD2 and rat CD48

Discussion

Acknowledgements

This project was carried out in collaboration with Dr. Neil Barclay and Dr. Simon Davis, The William Dunn School of Pathology, Oxford. CD5d1 was prepared by Dr. Mark McAllister and the CD5/CD4d3,4 chimera by Dr. Marion Brown. CD2 and sCD59 were prepared by Dr. Simon Davis

Abbreviations: WAX: weak anion exchange chromatography, RAAM: Reagent Array Analysis Method; 2AB: 2-aminobenzamide; CD: cluster differentiation; IgSF: immunoglobulin superfamily; ScR: scavenger receptor; V: variable; C: constant; GPI: glycosylphosphatidylinositol; NMR: nuclear magnetic resonance; N: neutral glycans; NA1, NA2, NA3, NA4: glycans with 1,2,3 or 4 negative charges respectively; CHO: Chinese Hamster Ovary; OGS: Oxford GlycoSystems: MALDI MS Matrix Assisted Laser Desorption Mass Spectrometry; P4 GPC: P4 Gel Permeation Chromatography; RAAM: Reagent Array Analysis Method; GC: gas chromatography; ANTS: sodium amino napthalene trisulphonic acid; FACE: Fluorophore Assisted Carbohydrate Electrophoresis; NDV; Newcastle Disease Virus; H: hexose, N; N-acetyl hexosaminidase; F: fucose

Abstract: The glycosylation of proteins is a complex event which is controlled by the interactions of the specific glycosylation machinery within the cell with each developing glycoform. In this chapter the extent to which the individual protein can control its final glycosylation pattern has been explored by analysing the glycosylation of the extracellular regions of the soluble recombinant forms of five leucocyte antigens all expressed in CHO cells. The antigens, which are members of different superfamilies, were human CD59 (Ly6), human CD2 (IgSF), human CD5 domain1 (ScR), human CD5/rat CD4d3,4 (IgSF) chimera and rat CD48 (IgSF). Sialylation varied from 100% in the chimera to 10% in CD2, oligomannose structures ranged from 46% in CD2 to 0% in CD5d1 and the chimera. CD59 contained the highest proportion of multiantennary and polylactosamine glycans (45%); at the other extreme CD5 contained none. Interestingly, CD5 contained only one glycoform and addition of the IgSF protein, CD4d3,4 to CD5d1 did not alter the processing significantly. Only CD5 and the chimera, which contain stretches of Pro/Ser/Thr contained O-glycans. These results indicate that both the developing glycan and the local protein structure around the glycosylation site are involved in the addition of monosaccharide residues by the tranferase enzymes. Although all glycoproteins begin in the endoplasmic reticulum with the identical oligosaccharide precursor, $Glc_3Man_9GlcNAc_2$, they emerge from the Golgi with a set of unique glycoforms whose processing is controlled by the protein to which they are attached.

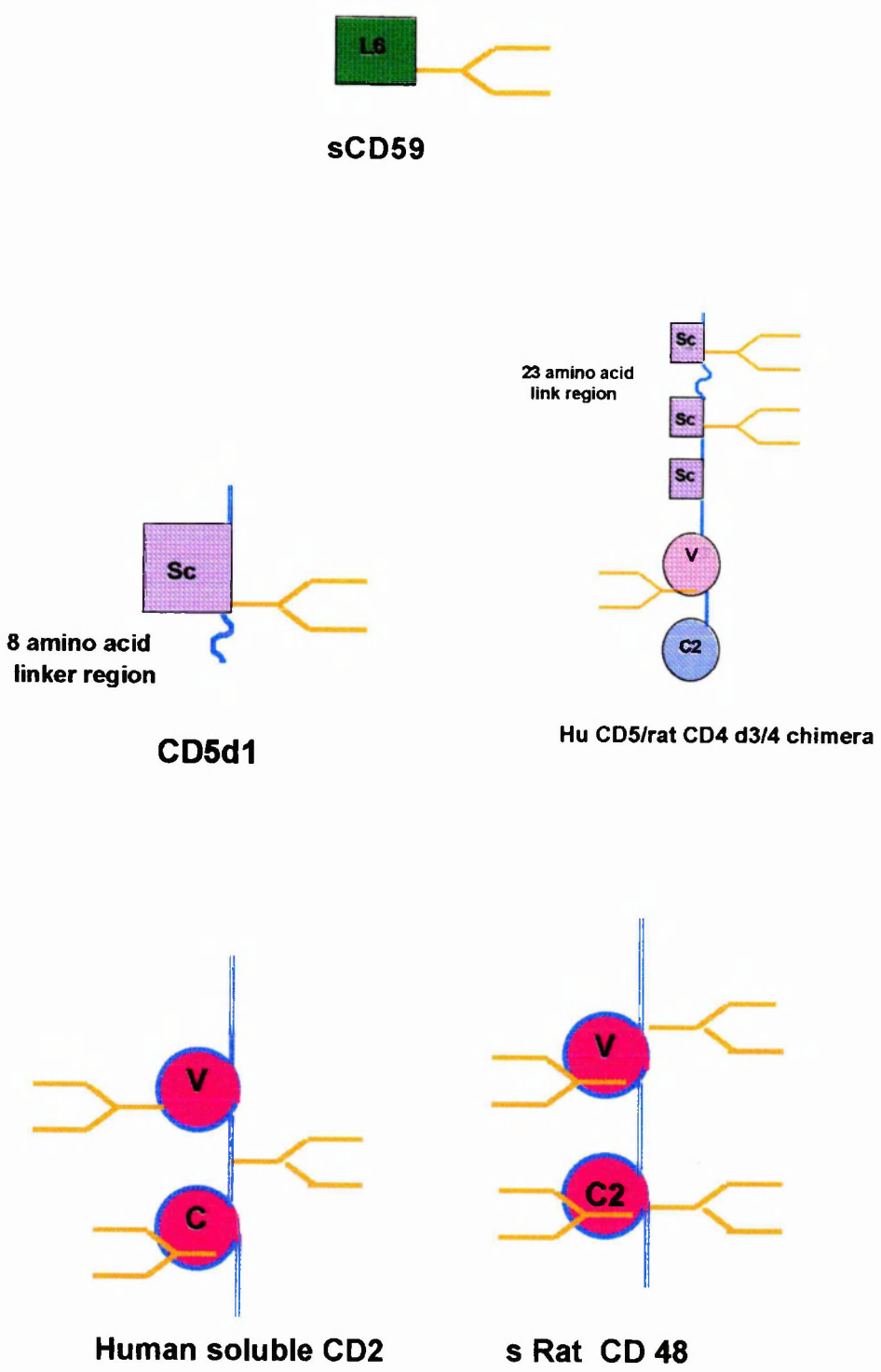


Figure 1: The glycoproteins which were analysed in this study.

Soluble human CD59, human CD5 domain 1, human CD5/rat CD4d3,4 chimera, human CD2 and human CD4 were all expressed in CHO cells.

Protein specific glycosylation patterns of five leucocyte cell surface antigens expressed in Chinese Hamster Ovary cells.

Introduction

1. Background:

1.1 T- cell surface antigens

It has been estimated that 20% of the T-cell surface is covered by characterised proteins many of which are heavily glycosylated (Williams et al 1987). These leucocyte antigens have been systematically named with a cluster differentiation (CD) number. The CD antigens can be grouped into approximately 10 different superfamilies according to sequence similarities in their extracellular domains. Three of these structurally distinct groups are the Ly-6, the scavenger receptor (ScR) families and the immunoglobulin (IgSF) superfamilies. In this chapter the glycosylation of a soluble form of a leucocyte antigen from each of these structural groups has been analysed (Fig.1). The extracellular regions of CD59 (Ly6), CD5 (ScR) domain 1, a chimera consisting of CD5 and CD4 (IgSF) domains 3 and 4, and CD2 (IgSF), were all expressed in the CHO-K1 cell line, allowing a study to be made of the extent to which individual proteins exposed to the same biochemical pathway can control their own glycosylation.

The Ly-6 family contains molecules with a structural unit consisting of 70-80 amino acids containing 10 cysteine residues. The protein was first identified in the mouse and subsequently homologues of the Ly-6 antigens have been found in rats but not in humans. Human CD59 has been identified as a member of the Ly-6 super family, however there are too many differences in the sequence for it to be a Ly-6 homologue. Interestingly, all the members of the Ly-6 family so far identified are attached to cell surfaces by a GPI anchor; in addition to the analysis of the glycans of soluble CD59 in this chapter, the glycans associated with the GPI anchor of human erythrocyte CD59 have been analysed in this thesis (ch3).

CD5, of which domain 1 was analysed in this study, was the first member of the ScR family to be identified. It contains three scavenger receptor domains in the extracellular region and these were later identified in other molecules including the macrophage ScR type I. The function of the C-terminal 110-amino acid cysteine rich domain is not known and no tertiary structural data is available. Interestingly, the class II macrophage scavenger receptor, which is identical to the class I type in every way except that the C-terminal domain is replaced by a 6-amino acid C-terminus, mediates the endocytosis of acetylated low density lipoproteins with essentially the same broad specificity and affinity as the type I molecules (Freeman et al 1990).

The IgSF is a large group of structurally related proteins that function in the immune system mainly in cell - cell recognition (Williams 1987; Williams and Barclay 1988). The various members of the superfamily are built of domains of approximately 100 residues each having a structure formed by two β-pleated sheets packed face to face. Individual members can differ in the number and size of strands in the two β-sheets, and in the size and conformation of the

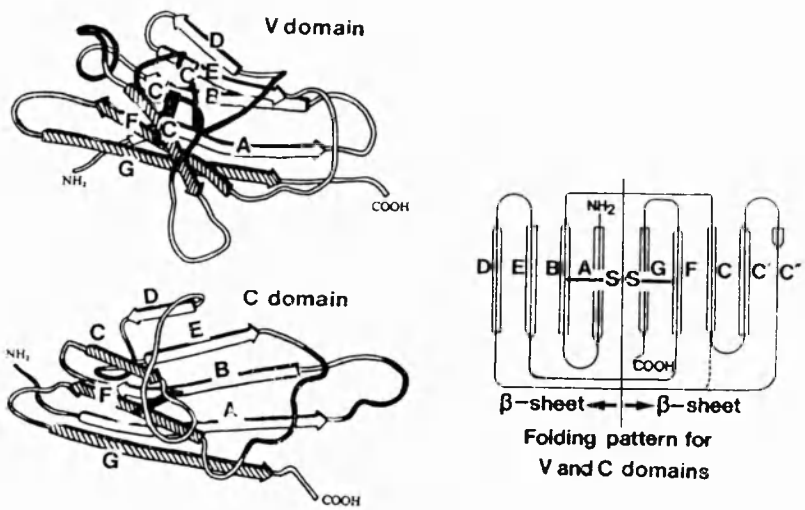


Figure 2: The patterns of the immunoglobulin fold. The labelling of the V_L and C_L domains is illustrated in the schematic diagram.

The basic structural component of the immunoglobulin superfamily is the immunoglobulin fold. The two β-sheets of the fold consist of anti-parallel β-strands containing 5-10 amino acids. Between the sheets a hydrophobic interior is formed from in-pointing hydrophobic amino acids that alternate in the β-strands with hydrophilic residues that point to the outside of the fold. The interaction between the sheets is stabilised by the conserved disulphide bond. The core of the fold consists of β-strands A, B and E in one sheet and G, F, and C in the other. These strands come from the first and last parts of the domain sequence, while in the middle there is considerable variation in sequence length. V and V-related domains have about 65-75 amino acid residues between the conserved disulphide bond, and there are 4 β-strands in each β-sheet plus a short β-strand segment across the top of the domain. In C domains the sequence between the disulphide bond is shorter (55-60 residues), giving rise to sheets with 4 and 3 β-strands.

(Taken from Williams and Barclay 1988.)

links between the strands. Nevertheless, similarities in sequences and structure allow the molecules to be grouped into sets based on three structurally distinct domains: variable (V), constant (C)1 and C2. The basic structural component of the immunoglobulin superfamily is the immunoglobulin fold (Fig.2). The Ig-related domains of non-Ig molecules are described as belonging to V- or C-sets depending on whether they have a pattern of strands approximating to a V or C domain. The C1 set are distinguished from C2 set in some conserved sequence patterns. The glycosylation of three members of the IgSF, which all contain V and C2 domains, have been studied in this chapter: human CD2, rat CD48 and rat CD4 domains 3 and 4 which were expressed as a chimera with human CD5 domain1.

1.2 The glycosylation of the leucocyte antigens in this study

Most of the leucocyte antigens are glycoproteins. The number of glycosylation sites varies from one molecule to another. For example, CD2 (M.Wt.37kD) has 3 Asn-linked sugars, CD5d1(11kD) has 1, the CD5/CD4,d3,4 chimera (76kD) has three, CD48 (26kD) has 4 and CD59 (19kD) has 1 (Fig.1). In most cases O-linked glycans have not been identified, although most molecules contain potential O glycosylation sites (Ser/Thr residues). Some, such as CD5, contain a peptide region rich in Ser, Thr and Pro residues and this is a common feature of O-glycosylated peptides (see ch.8 section 3 discussion). The approximate sizes of a V-domain and the glycans which are attached to it can be compared. The molecular weight of an N-linked sugar is of the order of 2kD, the length of a complex biantennary glycan is about 3nm, each sugar is about 0.6nm in length. A V domain is about 3nm long and has a molecular weight of about 12.5kD. On a weight for weight basis the sugars occupy a higher molecular volume than the proteins which are more densely packed, therefore in many instances the length of the sugar and the diameter of the antigen domain to which it is attached are of the same order of magnitude. The glycosylation of leucocyte antigens has not been analysed in detail, with the exception of rat and human soluble CD4 (Ashford et al 1993), rat CD2 (Davis et al 1993) and of one of the sites on CD2 (Recny et al 1992). The functions of the sugars are not generally known. The ligand (CD58 in human and CD48 in rat) and antibody binding properties of CD2 were reported to be dependent on glycosylation (Recny et al 192; Withka et al 1993), although this has not been confirmed (Davis et al 1995). However de-N-glycosylation of human erythrocyte CD59 with N-glycanase resulted in an 88% reduction of the C5b-9 complement inhibitory activity of the protein, and it has been proposed that the glycans attached to CD59 may orient the binding site towards its ligand, the C5b-9 complex (Ninomiya et al 1992).

1.3 The role of the protein in oligosaccharide processing

Some of the general principles which control glycosylation were discussed in chapter 1. In the following section some of ways in which the protein structure regulates the fine details of glycan processing are addressed.

The protein structure controls the specific glycosylation at individual sites

Within a specific cell type the protein structure may influence the class of glycan, the extent of elongation, and the termination steps in a site specific manner. A single, multiply glycosylated glycoprotein starting in the endoplasmic reticulum with identical glycans at each site may contain non-identical structures after processing. Although different glycan structures in different proteins from the same cell could be explained on the basis of different compartments and enzyme exposures, this is not the case for a single molecule, and this indicates that each of the initially identical oligosaccharide precursors in a protein represents a different structure as substrate for the processing enzymes. For example, in rat brain Thy-1 site 23 (Asn-Asn-Thr) contains only oligomannose structures while site 98 (Asn-Lys-Ser) is associated with complex, hybrid and oligomannose structures and site 74 (Asn-Phe-Thr) contains both complex and hybrid structures (Parekh et al 1987). The local three dimensional structure of the protein determines the proximity of the protein surface to the glycans and this may be an important factor which determines the glycosylation pattern at individual glycosylation sites on proteins expressed in a single cell by controlling the affinity of different enzymes for the developing glycoforms. In multidomain proteins the quaternary structure may also influence site specific glycosylation and this is demonstrated by a study of the MAC 1 integrin which contains the α^M subunit and LFA1 integrin which contains the α^L subunit. These glycoproteins have a common subunit (β) and the $\alpha\beta$-dimers form before processing making it possible to compare the glycosylation of the β subunit sites and to determine the influence of different α-subunits on processing (Dahms and Hart 1986).

The protein influences the class of glycans

In the early stages of processing the protein may shield the oligosaccharide from the mannosidase enzymes, so that in bovine pancreatic RNase B, 6% of the structures are not processed beyond Man 9, 17% remain as Man 8 glycoforms, 11% terminate at Man 7, 20% of the glycans reach the Man 6 stage and only 48% are processed fully by the mannosidase enzymes to Man 5. In other molecules such as IgG, no oligomannose structures are detected in the final population of glycoforms. Interestingly, the protein structure may also affect some of the later steps in glycan processing, when the terminal sugars are further from the protein, as well as the earlier steps when the smaller oligosaccharides are close to the protein surface. In a model system the relative rates of reaction were determined for individual glycosyl transferases presented with a single substrate which was either free or exposed at different distances from avidin protein (Yet et al 1988). Substrates were prepared in which glycans derivatised with asparaginyl biotin were immobilised to avidin with and without the insertion of a hexanoyl hydrocarbon spacer arm. The rates of reactions of the free biotinylated sugars were compared with those for the same sugars closely bound to the avidin surface or held at a distance by the hexanoyl spacer arm in assays in which each substrate was exposed sequentially to the same set of complex sugar processing enzymes. The major effect of the protein in the avidin derivatives was in promoting the early stages of glycosylation, especially the addition of GlcNAc to the Man 5 glycan. This step, which ensures that the sugars will be processed to complex and hybrid oligosaccharides, was essentially eliminated in the molecule which

complex and hybrid oligosaccharides, was essentially eliminated in the molecule which contained the spacer arm. Compared with the free biotinylated glycans the final termination steps were retarded in protein bound complex sugar populations which contained a greater proportion of the intermediate glycans, Gal1 and SAGal2. These data suggest that the distance of the protein surface from the oligosaccharides influences both the affinity of the GlcNAc tranferase and the extent of sialylation. In this thesis the rate of digestion of the α1,2 linked oligomannose sugars attached to RNase B with *A. Saitoi* α1,2 mannosidase indicated that this reaction was 150 times slower than the rate of digestion of the free sugars, and that the Man6 glycoform was more resistant than the Man7,8 or 9 (ch.4 section 4.1). This suggests that there may be a similar role for the protein in biosynthesis. The individual protein structure may also control the class of glycan by blocking key enzyme reactions and interestingly, in the study by Yet et al (1988) described above it was not possible to produce hybrid structures with the biotinylated substrates either when free or when avidin bound. Nor could tri- and tetra-antennary structures be developed, even on the free biotinylated glycans. Enzyme specificity stipulates a rigorous order of processing and these data suggest that if branching requires blockage of an alternative processing reaction, such as the addition of GlcNAc residues to the 1,3 arm, the protein structure is required to provide this. Alternatively, proteins with multiantennary glycoforms may contain positive effectors responsible for the additional branching suggesting that the protein may regulate the extent of branching not only by steric hindrance, but also through the stimulation of particular steps. Binding studies showed that the rate differences between the activities of the transferases with respect to the avidin bound and free substrates resulted from an altered turn over rate (Vmax) (typically 25x decrease in Vmax, 2-3x increase in Km going from free to bound substrate). Thus the processing GlcNAc transferases can recognise and bind either form of the substrate equally well but only 4% of the interactions are productive when the glycans are bound to avidin. It has been proposed that these non-productive enzyme-substrate complexes involve distortion of the conformation of either the enzyme or the substrate.

The conformation of the glycan may also influence the affinity of transferases for a particular glycoform and if, for example, the protein structure crowds the reducing end of the sugar the conformation of some glycoforms may be disturbed altering the affinity of the processing enzymes. In particular, the outer sugars on the 1,6 arm of oligomannose, hybrid or bisected complex sugars are flexible and can adopt a conformation in which the outer arm sugars are in close proximity to the reducing end (Brisson and Carver 1983). In some glycoproteins a particular conformation such as this may be inhibited by the proximity of the protein surface. In some instances the structure of the developing glycan may determine the structure of the final glycoform, for example, in the synthesis of sialyl Lewis X (sialylα2,3Galβ1,4(Fucα1-3)GlcNAc) α2,3 sialylation precedes outer arm α1,3 fucosylation, and the sialyl transferase is unable to process a fucosylated structure (Kitagawa and Paulson 1994; Lowe et al 1991). Both the glycan structure and the substrate specificity of individual enzymes may therefore impose rigid sequential order on the processing steps of branched oligosaccharides.

The protein influences the elongation and termination steps

The addition of sialic acid is a chain termination step. Consequently, in contrast to other monosaccharide residues, sialic acid is normally found only at the non-reducing termini of glycan chains, where it is linked α2- 3, -6 or, rarely, -8 to terminal galactose. Cells exert control over their glycosylation machinery to produce terminal glycosylation sequences in a cell type specific and developmentally regulated manner. CHO cells produce N-linked sugar chains with the terminal NeuNAcα2,3Galβ1,4GlcNAc-R sequence and are therefore believed to lack the Galα2,6-sialyl transferase which synthesises the 2,6 linkage. Consistent with this conclusion an α2,6 sialyl transferase transfected into CHO cells competed with wild type α2,3 sialyl transferase and glycans containing both 2,3 and 2,6 sialic acid were synthesised (Lee et al 1989). Interestingly, the subcellular location of the 2,6-sialyl transferase was similar to that of the wild type enzyme and extended from the cis to the trans Golgi indicating that termination can take place at many locations in the Golgi or ER, and as early as in the cis Golgi cisternae. *In vitro* studies with purified sialyltransferases have also demonstrated that Galα2,6-sialyl transferase competes with Galα2,3-sialyl transferase (Weinstein et al 1982). The role of the oligosaccharide in the addition of sialic acid to complex glycans has been explored in studies of several glycoproteins. The relationship between chain length, charge and sialic acid linkage has been investigated in glycoproteins from a mouse lymphoma cell line (Merkle and Cummings 1987), thyroid glycoprotein (Edge and Spiro 1985), and fetal lactosaminoglycan (Fukuda et al 1984). In general, these studies show that sialylation is more likely to occur in the early stages of processing and the incidence falls off as the chains become more elongated. Moreover, the sialic acid present on shorter oligosaccharides (less than 3 linear units of the repeating disaccharide 3Galβ1,4GlcNAcβ1) is more often α2,6 while the α2,3 linkage is more common in the relatively less sialylated longer polylactosamine extensions (Merkle and Cummings 1987). This suggests that the length of the glycan chain, and therefore the proximity of the protein matrix, are factors which regulate the addition of an α-2,3 or an α–2,6 linked sialic acid residue. However, a developing glycan containing terminal galactose need not necessarily be sialylated, but may terminate with galactose or be further elongated by the addition of terminal GlcNAc or polylactosamines. This suggests that in a homogeneous population of glycoforms the proportion taking each of the alternative processing pathways is a competitive event, controlled by the location and concentration of the various enzymes, the affinity of the developing glycoform for each and the transit time of each glycoform through the ER and Golgi.

1.4 Poly-N-acetyllactosamine extensions

Complex glycans which have not been terminated by sialic acid can be further extended by the addition of GlcNAc and Gal to the terminal galactose residues. The level of β-1,3 N-acetyl glucosaminyl transferase is the limiting factor for lactosamine (Galβ1,4GlcNAcβ1,3) extensions rather than β-1,4-galactosyl transferase and the terminal sugar for complex glycans carrying lactosamine extensions is normally galactose. Polylactosamine extensions were first identified on human erythrocytes which may express glycoproteins with 5-15 repeats in one side chain.

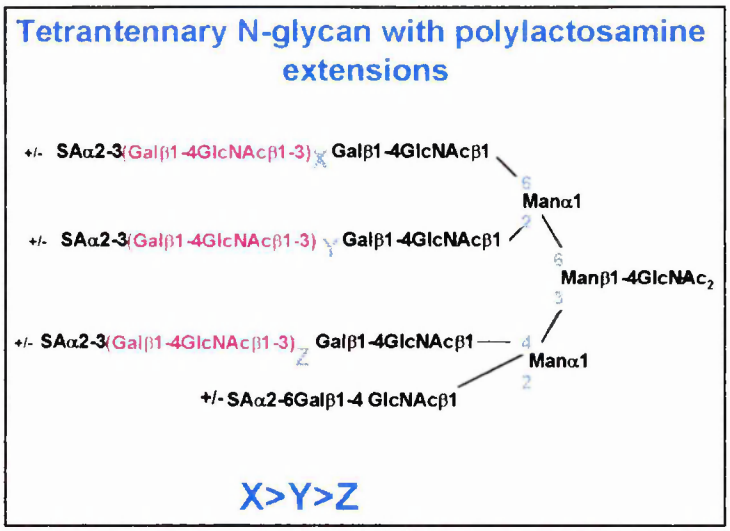


Figure 3: Complex tetrantennary structure with polyN-acetyllactosamine extensions *on three branches. The lactosamine extensions (x,y,z) are preferentially added to the α 1,6 arm of the Manα1,6 branch of the primary mannose (x) which is>y>z. Sialic acid is more commonly linked in the α2,3 position to the Manα1,6 arm of the primary mannose, while the α1,2 branch of the α1,3 arm is more often associated with α2,6 linked sialic acid.*

Both linear and branched polylactosamines may be modified to carry the ABO blood group antigens, and they also constitute the developmentally regulated I/i antigens. In infants under one year the I/i antigens are mainly linear polylactosamines (I), after which the structures become predominantly branched (i). Poly-N-acetyllactosaminyl repeats are more often found in the side chains attached to the α-mannose residue attached to the 1-6 arm of the C6β-mannose (fig. 3). Many of the glycoproteins which carry lactosamine extensions are membrane proteins. These include the type III membrane proteins, human erythrocyte band III (Fukuda et al 1984) and the lysosomal glycoprotein CD63 (Fukuda 1991). Based on these findings it has been suggested (Fukuda 1994) that the transmembrane section of the protein must cross the membrane more than once for lactosamine structures to be added to the glycans. However, thymocyte Thy-1, which has a glycosylphosphatidylinositol anchor contains 20% of polylactosamine structures (Parekh et al 1987) and in this thesis polylactosamines were identified on human erythrocyte CD59, which also contains a GPI anchor (ch3) rather than a transmembrane section. During processing proteins containing membrane anchors are located in the Golgi membrane and as a result may have extended transit times during which long polylactosamine extensions may be synthesised. Increased transit times have been shown to increase the polylactosamine content of the glycans attached to lamp proteins-1 and -2 (Wang et al 1991). In addition, the close proximity of anchored glycoproteins to the glycosylating enzymes, which are also anchored in the Golgi membranes, may increase the efficiency of the enzymes. The role of the protein in controlling the extension of the glycan chain has been investigated. In this respect the amino acid sequences of proteins carrying polylactosamines have been compared, but no consensus sequence has emerged that could provide a signal. However, it has been noted that the glycosyl transferases require access to the protein surface to process complex glycans. In the same way the three dimensional structure of the protein, including modifications such as O-glycosylation, may be important for the activity of β-1,3 N-acetyl glucosaminyl and β-1,4-galactosyl transferase enzymes in the processing of lactosamine extensions.

Thus the processing steps leading to complex or hybrid structures, extent of branching, and specific structures at the non-reducing terminals are determined by the interaction of individual oligosaccharides in the context of the local protein structure with the exquisitely specific processing apparatus available in each cell. In this chapter the possibility that differences in the structures of individual proteins may lead to oligosaccharide diversity has been explored in five leucocyte antigens expressed in the same CHO cell line.

2. The importance of controlling the glycosylation of recombinant leucocyte antigens

Many of the known functions of the glycosylated T-cell surface antigens involve protein-protein recognition and, for the most part, the roles of the oligosaccharides are not known. To probe aspects of recognition and function a number of cell lines have been established in which both intact proteins and constructs can conveniently be expressed, since the abundance of the

intact proteins and constructs can conveniently be expressed, since the abundance of the natural proteins is generally low in biological samples. It is well established that glycosylation is cell type specific (Parekh et al 1989) and that the activity of recombinant glycoproteins may, as has been demonstrated with tPA (Parekh et al 1989), be significantly modified by glycosylation. Cell specific glycosylation is therefore a factor to be taken into account when selecting cell lines in which to express recombinant glycoproteins. In the absence of information concerning the role of the sugars, it is important to maintain the integrity of the oligosaccharides and to avoid problems with events such as folding, transport or secretion which may depend on specific glycan structures. This may be achieved by expressing the recombinant proteins in cell lines where the biosynthetic pathways are similar to those of the cell in which the natural protein is expressed. The Chinese Hamster Ovary cell line produces stable cells which are readily grown and transplanted; this cell line also has glycosylation machinery competent to modify proteins with a wide range oligosaccharides including those commonly associated with human glycoproteins.

Functional studies may include an examination of the role of glycosylation, and one common approach is to establish whether the N-linked oligosaccharides are essential for a particular function by comparing the activity of the natural material with aglycosylated or deglycosylated protein. There are several methods by which non-glycosylated protein can be prepared. First, it may be possible to de-N-glycosylate the native protein with PNGase F, and as has been shown in this thesis (ch.3 and 4); this depends on the extent to which the local protein structure can shield the sugar-amide linkage. An alternative method is to grow cells in the presence of tunicamycin, which inhibits the addition of the oligosaccharide precursor, $GlcNAc_2Man_9Glc_3$, so that aglycosylated proteins are expressed. Cells grown in the presence of N-butyl deoxynijiromycin (DNJ), which inhibits the removal of glucose from the precursor, are secreted only as Man_9Glc_3 glycoforms from which sugars can be removed by endo H (Davis et al 1995) which cleaves within the chitobiose core. There is also the possibility of making significant mutations such as replacing the asparagine residue required for the addition of the precursor to the protein although this involves a change in the primary sequence of the protein which may have important implications for the structure. A powerful approach which has been developed recently is to develop strains of animals or cells in which the genes for selected glycosylating enzymes have been deleted (Stanley et al 1994, Metzler et al 1994).

Analysis of the glycans is an important complement to the structural analysis of glycoproteins, since this is normally based on techniques such as X-ray crystallography and nuclear magnetic resonance (NMR), which do not define the glycosylation in detail.

3. Introduction to the studies in chapter 2

In this chapter the glycans attached to CD2 and soluble CD59 were characterised to give further information to the NMR and X-ray crystallography studies of these samples. Non-glycosylated human erythrocyte CD59 showed an 88% reduction in complement inhibitory

In addition to determining glycosylation to complement structural and functional analysis this study also explores the extent to which an individual glycoprotein can control the biosynthetic processes which determine its own glycosylation, subject to the biosynthetic pathways available in a cell. The first stage in the N-linked glycosylation process involves the transfer of a $Glc_3Man_9GlcNAc_2$ oligosaccharide precursor to some asparagine residues in the protein. During transport of the glycoprotein through the ER and Golgi this common precursor may be processed in different ways by the glycosidase and glycosyl transferase enzymes which are located in the ER and the Golgi stacks. A single glycosylated variant (glycoform) of a protein is the product of the particular series of enzyme reactions which have processed the $Glc_3Man_9GlcNAc_2$ structures which were initially attached to it, and in general glycoproteins consist of collections of glycoforms. Within the same cell the glycosylation profiles of the oligosaccharides attached both to different proteins and to different sites on the same protein may be distinct from each other, although all the glycoproteins have been exposed to the same glycosylation machinery. This suggests that the local protein structure close to the glycosylation sites has some control over the sequence of enzyme reactions which determine its glycosylation. In this study the sugars attached to five cell surface antigens expressed in the same Chinese Hamster Ovary (CHO) cell line were analysed to investigate the extent to which individual glycoproteins influence the biosynthetic processes which determine the structure of their own glycans. The antigens are discussed in three sections: (A) soluble human CD59, (B) human CD5 domain 1 and a chimera consisting of human CD5 and rat CD4 domains 3 and 4 (C) human CD2 and human CD48. In a complementary study (ch3) the glycosylation of human erythrocyte CD59 was analysed.

4. Methods

The general strategy shown below was developed for the analysis of the glycans attached to leucocyte antigens and some approaches to oligosaccharide sequencing involving newly developed anion exchange chromatography and mass spectrometric techniques were explored. (Details of these are in ch.8, which also includes the oligosaccharide sequencing methods used throughout this thesis.)

1. Release of glycans by hydrazine at $95^{o}C$
2. 2AB and tritium labelling
3. Determination of relative amounts of N, NA1 and NA2 glycans by WAX chromatography
4. Determination of nature of charged species: digestion with *Arthrobacter Ureafaciens*/ Newcastle disease virus neuraminidase
5. Separation of asialo glycans by P4 GPC
6. Analysis of oligomannose series by exoglycosidase digestions analysed by P4 GPC and MS
7. Analysis of asialo N-linked glycan pool by exoglycosidase digestions analysed by MALDI MS
8. Analysis of purified individual asialo N-linked glycans by RAAM analysed by P4

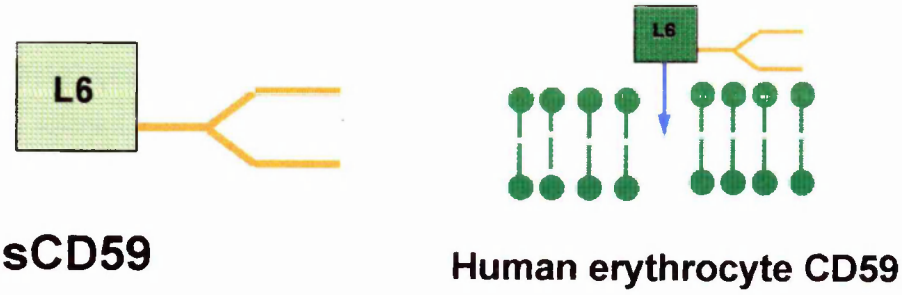


Figure1a: A schematic representation of CD59. *The glycosylation of soluble (s) CD59 expressed in CHO cells is discussed analysed in this chapter. The glycosylation of the native form of the molecule, which contains a GPI anchor, is discussed in chapter 3.*

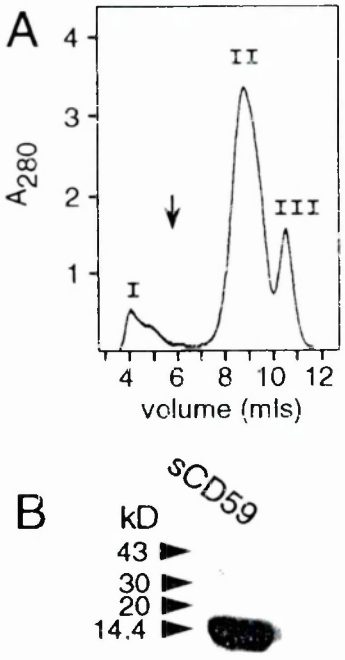


Figure 1b: Preparation of recombinant sCD59 expressed in CHO cells. (A) sCD59 purified by affinity chromatography was applied to a Superdex S-75 gel filtration column and eluted in three fractions (I,II and III). The arrow indicates the elution position of CD2 used as a molecular weight marker. (B) 5mg of fraction II used for oligosaccharide analysis was subjected to SDS PAGE analysis under reducing conditions in a 20% gel and stained with Coomassie blue (Kieffer et al 1994).

Ch. 2A The glycosylation of soluble human CD59 expressed in CHO cells

1. Background

The extracellular region of CD59 (Fig.1a) consists of a single Ly-6 superfamily domain containing two N-linked glycosylation sequons (Asn8ProThr and Asn18CysSer), one of which (Asn8) is not occupied. CD59 is attached by a glycosylphosphatidyl inositol (GPI) membrane anchor to the cell surfaces of a wide variety of cell types including leucocytes, endothelial and epithelial cells, erythrocytes and placenta where it restricts the cytolytic activity of homologous complement by binding to C8 and C9 (Walsh et al 1992, Lachman 1991). sCD59 is a relatively flat, disc shaped molecule consisting of a two stranded β-sheet finger loosely packed against a protein core formed by a three stranded β-sheet and a short helix (Kieffer et al 1994). The structure and function of human erythroctyte CD59 is discussed in detail in the background to chapter 3.

Significant quantities (up to 2mg/l) of the extracellular region of CD59 required for NMR structural analysis (Kieffer et al 1994) were produced (Dr. Simon Davis, William Dunn School) by heterologous expression of the truncated cDNA in CHO cells. CD59 cDNA was modified by the insertion of a stop codon after the codon for Asn-70 of the mature protein sequence. This eliminated the region coding for carboxy-terminal segment which contains a cluster of 18 mainly hydrophobic amino acids (residues 87-105) preceded by 8 polar amino acids, which, together define a typical signal peptide for GPI anchored proteins. The cleavage site for the anchor attachment, which is between Asn 77 and Gly78, (Tomita et al 1991) was also deleted. This soluble form of CD59 was purified from the supernatant of cell cultures of CHO cells by affinity chromatography using YTH53.1 monoclonal antibody (Davis et al 1989) and gel filtration. SDS PAGE analysis indicated that fractions II and III from the gel filtration column contained sCD59. Fraction III contains unglycosylated CD59, suggesting that glycosylation is not required for folding or secretion. The molecular weight of the glycosylated fraction II was approximately 14kD (Fig. 1b). NMR studies of this material identified resonances of the N-acetyl groups from at least three N-acetyl glucosamine residues, including two forming part of the pentasaccharide core common to N-linked glycans. The absence of a network of glycan-protein interactions suggested that the oligosaccharides behave as a flexible appendage to the protein. The analysis of the glycosylation, which is the subject of this chapter, was undertaken to provide additional structural information to complement the NMR analysis and also to compare the glycosylation of sCD59 with that of other proteins expressed in CHO cells to explore the effect of the 3-dimensional structure of the protein on glycosylation patterns. The oligosaccharide sequencing methods used here are described in chapter 8 of this thesis, and some details of the function of CD59 in controlling homologous lysis are discussed in chapter 3.

2. Strategy

Glycans were released from sCD59 using the OGS GlycoPrep 1000. Fractions of the glycan pool were radiolabelled with tritium, fluorescently labelled with 2-aminobenzamide (2AB) or retained unlabelled for MS analysis. The relative amounts of neutral glycans (N), N-linked glycans containing a single charged group (NA1) and N-linked glycans with 2 charged groups (NA2) were determined by weak anion exchange (WAX) chromatography. The linkages and type of the charged species were characterised by determining the susceptibility to *Arthrobacter Ureafaciens* and Newcastle Disease Virus sialidases. Hydrodynamic volumes in glucose units (gu) were determined using P4 GPC. The molecular weights of the components of the glycan pool were determined by MALDI MS, and the products of exoglycosidase sequencing were monitored by MALDI MS. The pool was separated by charge and analysed by simultaneous exoglycosidase sequencing and mass spectrometry. A second fraction of the pool was desialylated and the components analysed by P4 GPC and exoglycosidase digestions involving the Reagent Array Analysis Method (described in chapter 8 secton 1.8).

3. Results

Glycans (130nmoles) were released from sCD59 (1.9ng) using the OGS GlycoPrep 1000 (The manufacturers estimate the recovery is >85%). 80% was radiolabelled with tritium (total counts 7.2E8). 2% was labelled with 2AB; the remainder was retained for MS analysis.

3.1 Charge analysis of glycans released from sCD59

A fraction of the released glycans (2%, 2.6nmoles, 3.6E6) was analysed by WAX chromatography (Fig.2a) and the peaks assigned by reference to a set of fetuin standards resolved under the same conditions (Fig.2b). The glycans were again analysed by WAX chromatography after digestion with *Arthrobacter ureafaciens* sialidase (Fig.2c) and with Newcastle Disease Virus sialidase (Fig.2d). All structures are reduced to neutral with both sialidases, indicating that the charge is due to 2,3-linked sialic acid.

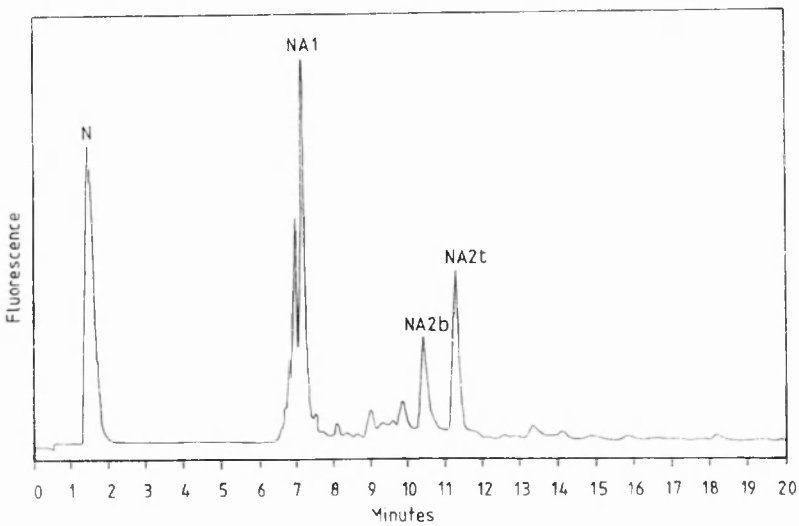


Figure 2a: WAX chromatography profile of CD59 2AB labelled glycan pool.
N (neutral): 35%; NA1 (monosialylated): 35%; NA2b (disialylated, biantennary) 12%; NA2t (disialylated, tri-antennary or polylactosamine) 18%.

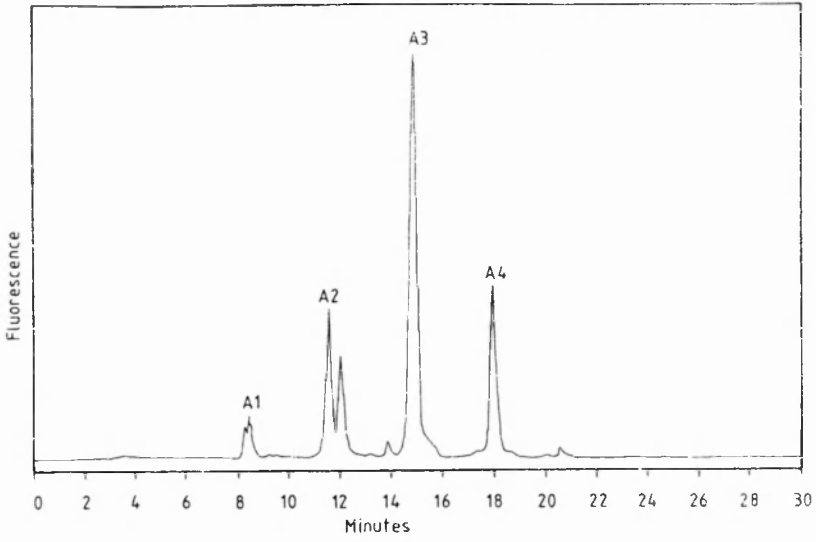


Figure 2b: Standard fetuin N-linked glycans analysed by WAX chromatography

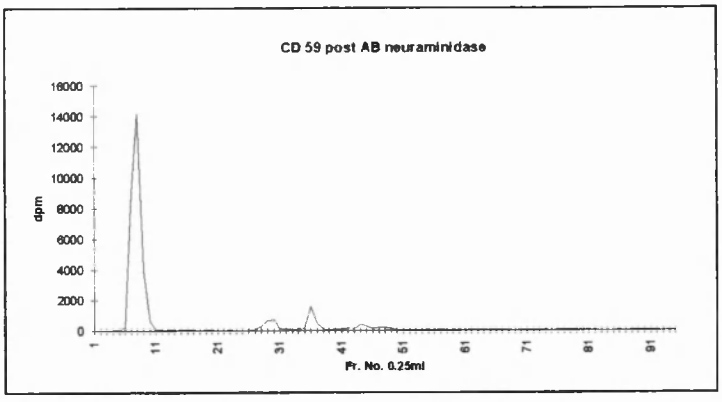


Figure 2c: WAX chromatography profile of tritium labelled CD59 glycans post *Arthrobacter ureafaciens* neuraminidase digestion.

All charged species have become neutral; all charge is due to α2,3 or α2,6-linked sialic acid.

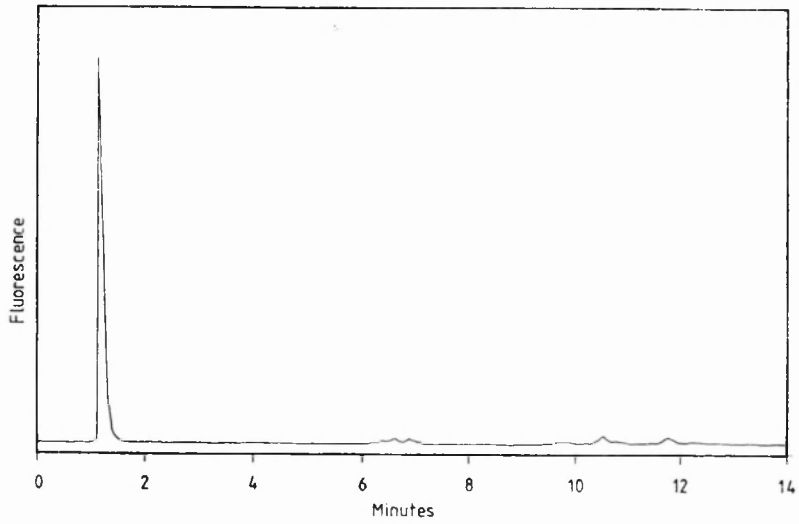


Figure 2d: WAX chromatography profile of 2AB labelled CD59 glycans post Newcastle Disease virus neuraminidase digestion.

All charged species have become neutral therefore all charge is due to 2,3 linked sialic acid.

Conclusion:

Charge profile of glycan pool released from sCD59:
N: 35% NA1: 35% NA2: 30%

3.2 Analysis of asialo glycans from sCD2 by hydrodynamic volume using P4 GPC

A fraction of the tritium labelled glycan pool (53%; 70nmoles; 4E4cpm) was de-sialylated with *Arthrobacter ureafaciens* sialidase and analysed by P4 GPC (Fig. 3, Table 1) and MALDI MS (Fig. 4, Table 2). Schematic representations of the structures are shown in figure 7.

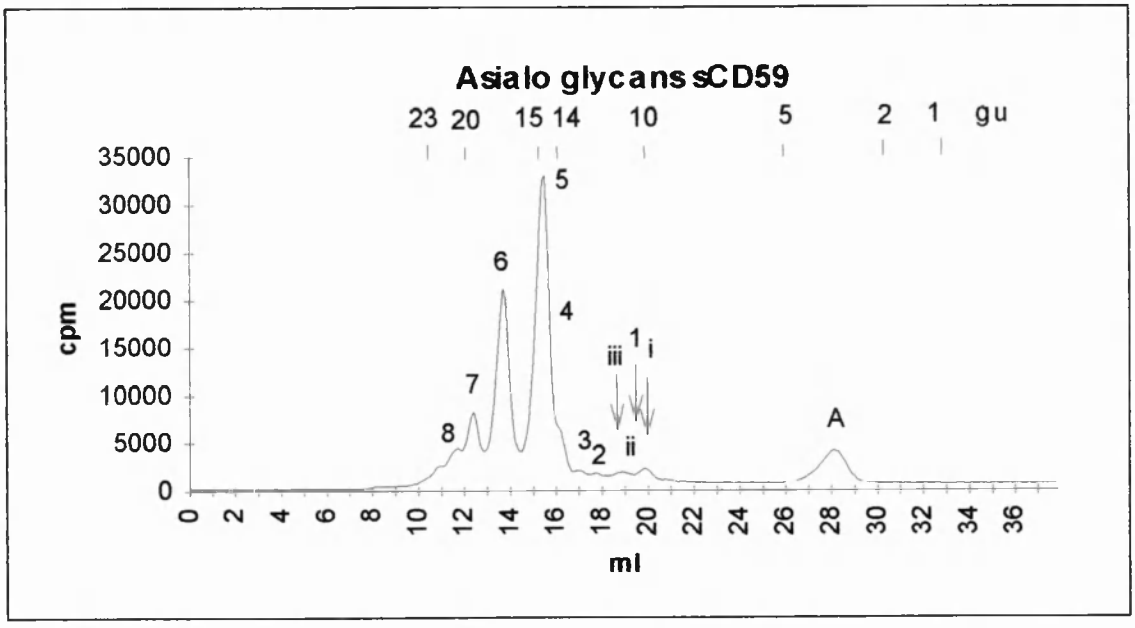


Figure 3: P4 GPC profile of asialo glycans from soluble human CD59

Peaks 1-8 contain complex glycans, peaks I-III contain oligomannose structures (Table 1 and figure 7).

3.3 Structural analysis of asialo glycan pool by sequential exoglycosidase digestions analysed by MALDI MS

The asialo glycan pool was sequentially digested with the following enzymes:

1. Bovine Testes β-galactosidase
2. Jack Bean β-hexoseaminidase
3. Bovine Testes β-galactosidase
4. Jack Bean β-hexoseaminidase
5. Choronia Lampas α-fucosidase

After purification on mixed bed resins the products were analysed by MALDI MS (Fig. 5b-f).

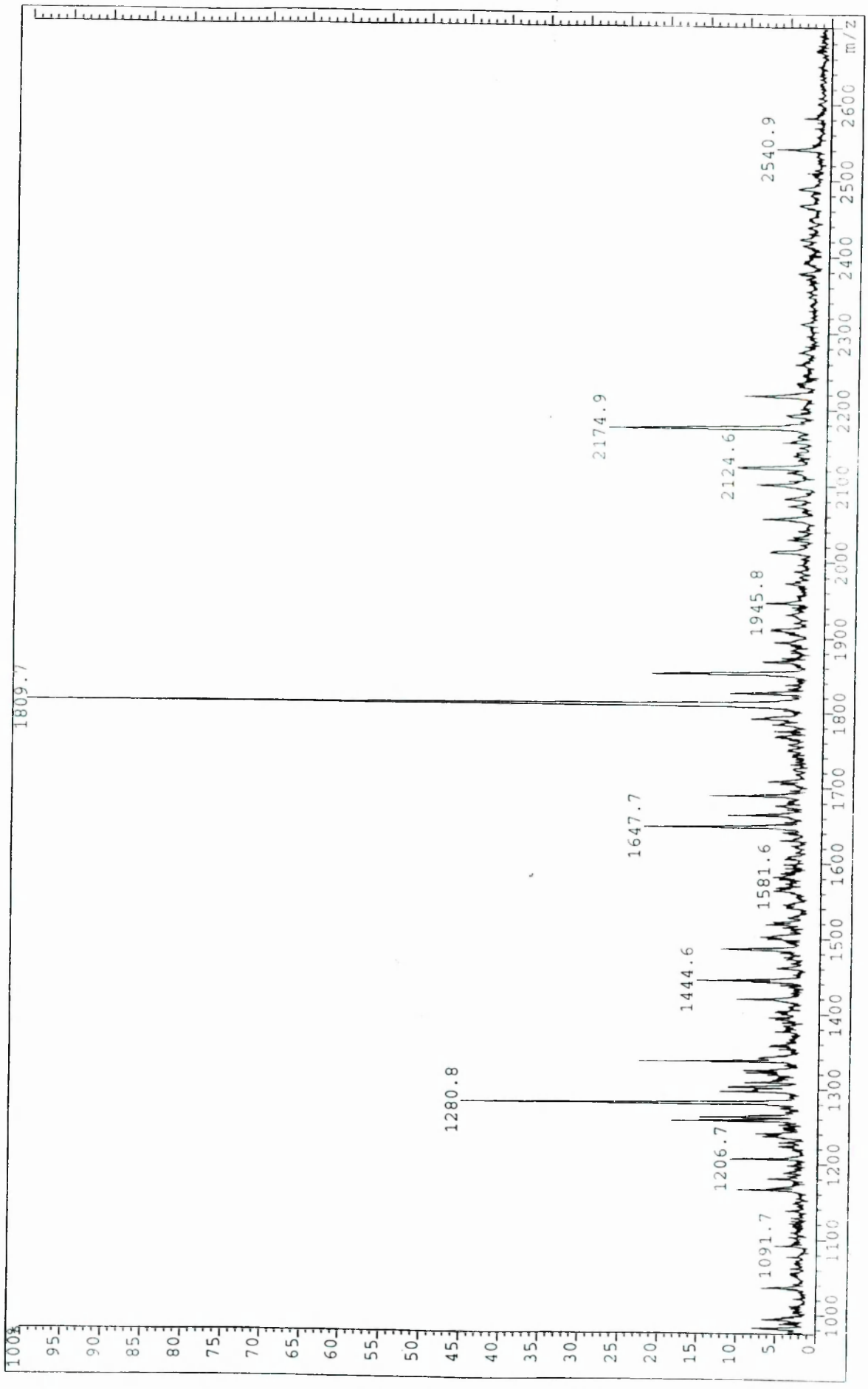


Figure 4: *MALDI MS analysis of asialo glycans from sCD59. The data in this figure are analysed in Table 2 columns 2 and 3 and in figure 5.*

	Peak no.	G.U.	%
M5	i	9.1	2.8
H3N3F	1	10	1.1
M6	ii	10.2	1.4
M7	iii	11.4	1.7
H4N3F	2	11.5s	
H3N4F	3	12.5	1.8
H4N4F	4	13.5s	
H5N4F	5	14.7	46
H6N5F	6	18	28
H7N6F (t)	7	21.2 (t)	9.2
H7N6F (p)	7	21.2 (p)	
H8N7F?	8	23.6	4.8
?		26.4	3.2

Table 1: Analysis of data from figure 3, the P4 GPC of sCD59

M: mannose; H: hexose; N: hexosaminidase; F: fucose

	Ch. mass	MS : m/z	BT Gal	JB Hex	BT Gal(2)	JB Hex(2)	CL Fuc
M5	1258.1	1257.9	1257.9	1257.4	1257.8	1257.8	1259.2
H3N3F	1283.13	1280.8	1282.5	1079.4	1080.2	1079.4	934.5
M6	1420.24	1418.9	1419.5	1419.8	1419.9		1419.8
M7	1582.38	1581.6	1485.6				
H4N3F	1445.27	1444.6	1282.5	1079.4	1080.2	1079.4	934.5
H3N4F	1486.32	1486.6	1485.6	1079.4	1080.2	1079.4	934.5
H4N4F	1648.46	1647.7	1485.6	1079.4	1080.2	1079.4	934.5
H5N4F	1810.6	1809.7	1485.6	1079.4	1080.2	1079.4	934.5
H6N5F	2175.93	2174.9	1850.6	1444.8	1283.5	1079.4	934.5
H7N6F(p)	2541.26	2540.9	2215.8	1809.9	1486.8	1079.4	934.5
H7N6F (t)	2541.26	2540.9	2053.8	1444.8	1283.5	1079.4	934.5

Table 2: Analysis of data from figure 4, the MALDI MS of sCD59 (columns 1-3) and of the sequential exoglycosidase digestions of the sCD59 asialo glycans (columns 3-7).

Ch. mass : chemical mass; MS:m/z : experimental molecular weights from the m/z values;

BT Gal: bovine testes galactosidase; JB Hex: Jack Bean β-hexosaminidase; CL Fuc : *Choronia Lampas* fucosidase.

For diagrams of structures see figure 7.

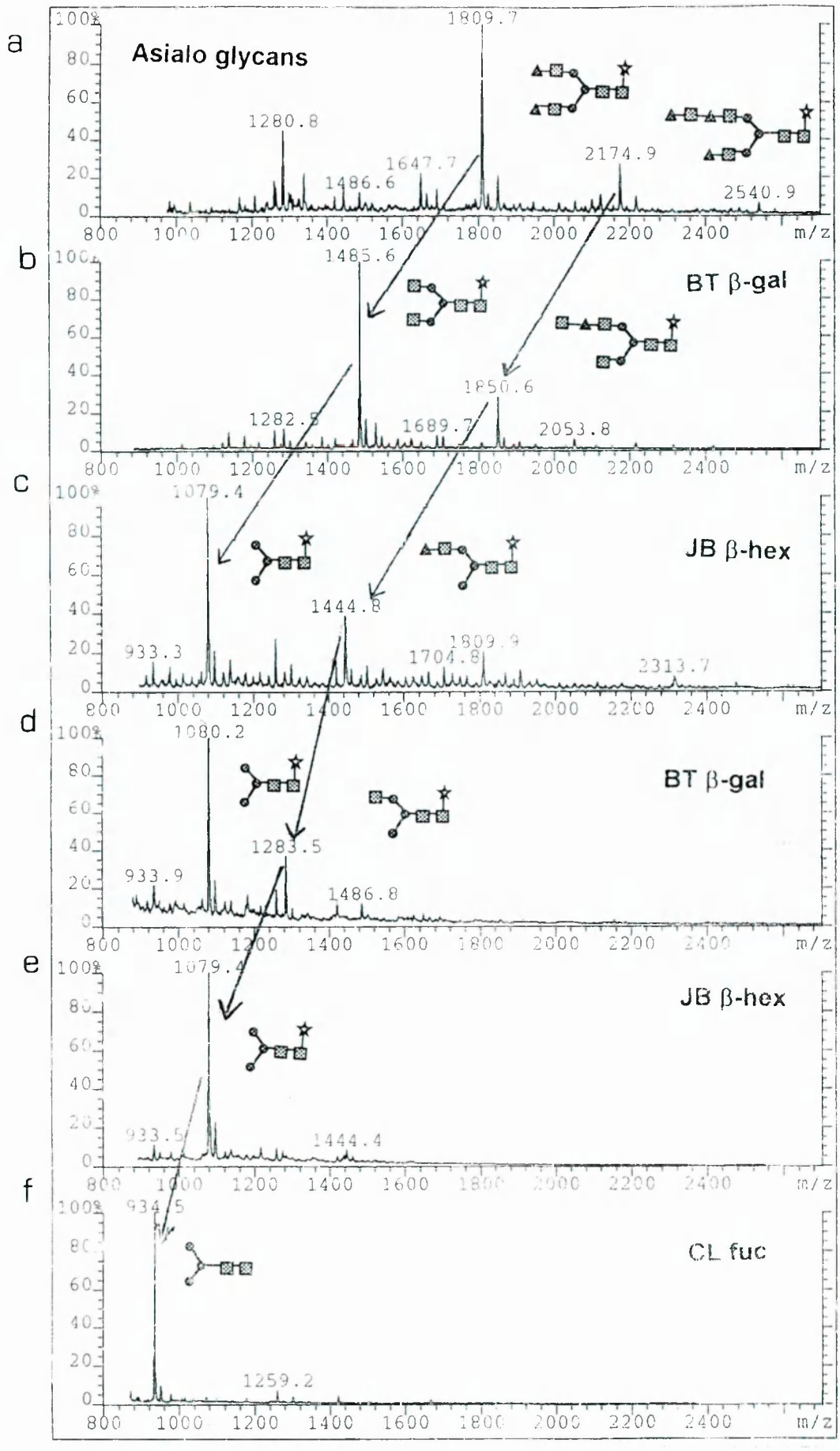


Figure 5a-f: MALDI MS analyses of sequential enzyme digestions of sCD59 asialo glycans. *(a) pool of asialo glycans from CD59 (b) after digestion with Bovine Testes β-galactosidase (c) after Jack Bean β-hexoseaminidase (d) after Bovine Testes β-galactosidase (e) after Jack Bean β-hexoseaminidase (f) after* Choronia Lampas *α-fucosidase. The results of this analysis are summarised in Table 2, and the information is incorporated into figure 7.*

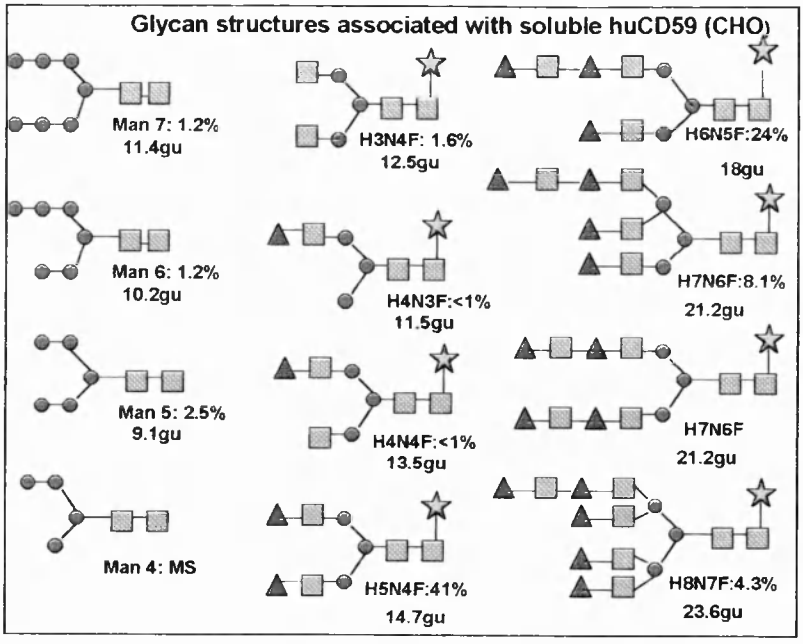


Figure 7: Asialo glycans from sCD59 *analysed by a combination of exoglycosidase sequencing, P4 GPC and MALDI MS showing termination of glycosylation at many points in the biosynthetic pathway. MS does not give linkage or arm specificities. Although some of this information was obtained by the use of specific exoglycosidase enzymes, the arm specificity assigned to the triantennary glycans is based on the probability that the GlcNAcα1,6Manα1,6 arm is the most likely to be extended (Fukuda et al 1984).*

3.4 Confirmation of the structures of the two major asialo glycans using RAAM

The 14.4 and 17.5gu peaks from figure 3 were identified as H5N4F and H6N5F as shown in figure 6. Peak 7 (21.2gu) was also subjected to RAAM analysis. No matching signatures were found, possibly because there are both tri- antennary and biantennary glycans with the same structural formula within this peak.

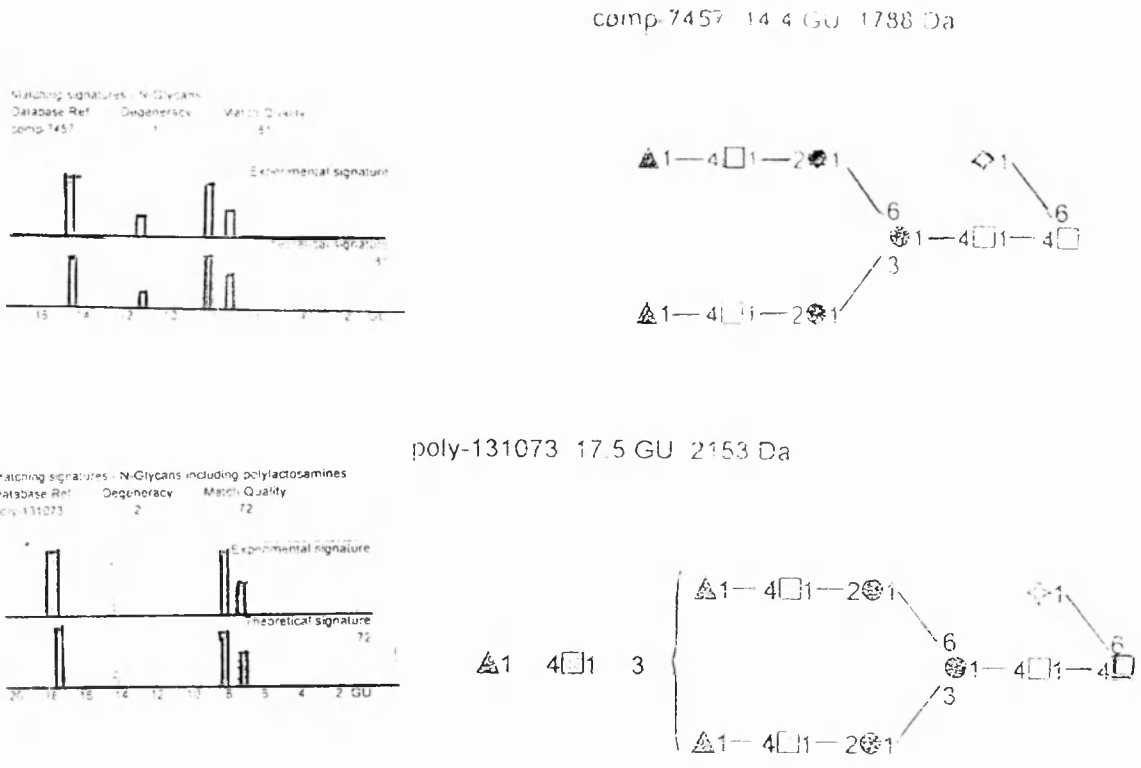


Figure 6: RAAM analysis of peaks 5 and 6 from figure 3

3.5 Analysis of peak A (3gu). Attempts were made to identify this by MALDI MS. No signals were obtained at m/z ratios which correlated with known glycans and may be an artefact arising from an environmental contaminant. The peak was resistant to Bovine Testes β-galactosidase.

A combination of sequential and RAAM exoglycosidase sequencing, P4 GPC and MALDI MS analysis allowed the structures in figure 7 to be assigned to the asialo glycans of sCD59.

3.6 Analysis of CD59 glycans after fractionation by charge

The 2AB labelled fractions from each of the peaks N, NA1, NA2b and NA2tp (Fig.2a) were desialylated with *Arthrobacter ureafaciens* neuraminidase and analysed by MALDI MS (Fig.8a-d). The structural formulae consistent with the molecular masses, including the 2AB label (120D), are summarised in Table 3. All the glycans previously detected were present as neutral structures. All the complex glycans from H5N4 and above were also present both as mono- and di- sialylated species. In addition a small percentage of H5N5F was detected in the NA1 fraction. This had not been identified in the earlier analyses.

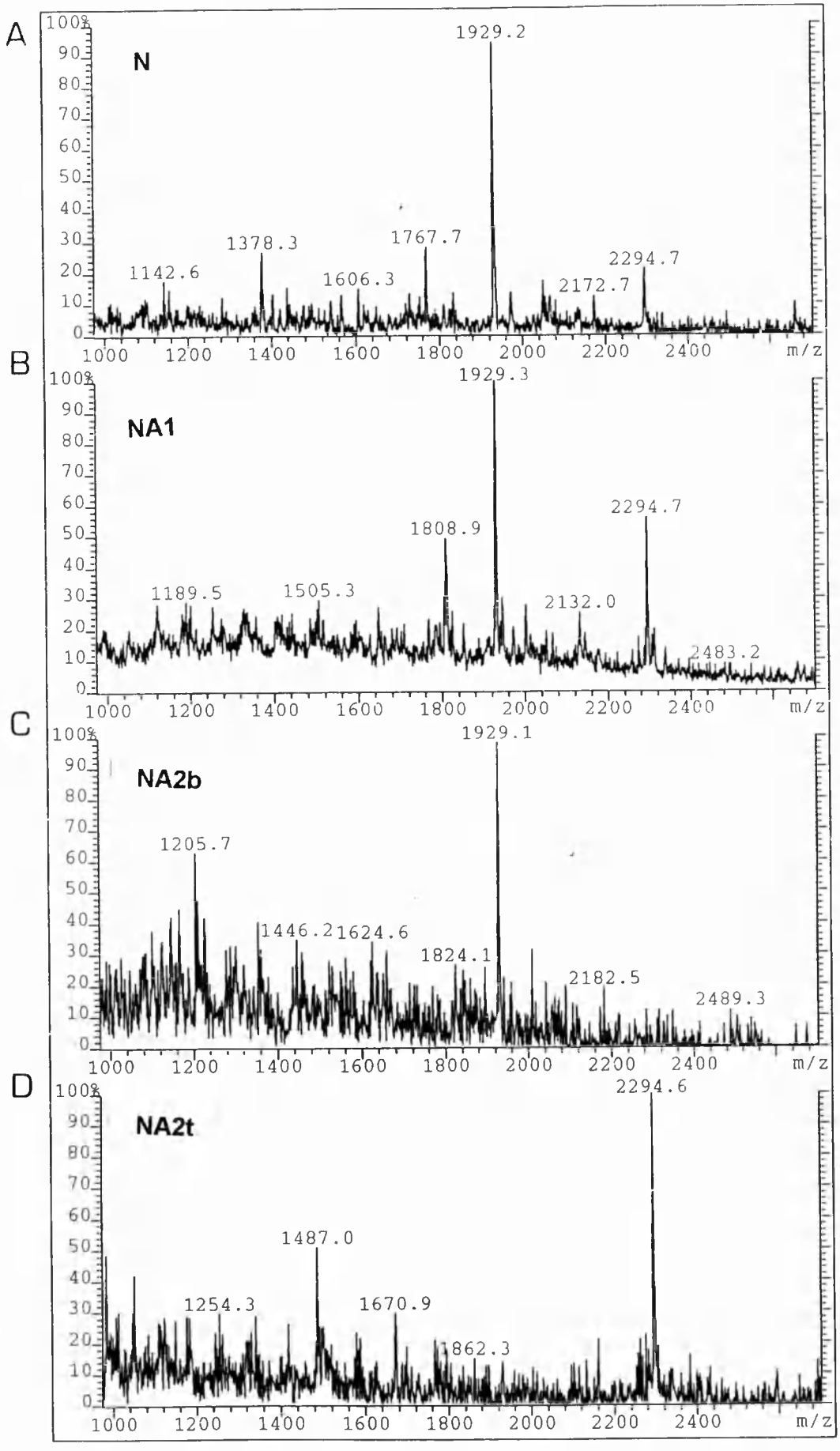


Figure 8: The MALDI MS analysis of asialo 2AB labelled glycans from peaks N, NA1, NA2b and NA2tp (figure 2a). The data are analysed in table 3.

	N(O)	N(C)	NA1	NA2(b)	NA2(tp)
M5	1378.3				
M6	1540.4				
M7	1702.4				
H3N3F		1280.2			
H4N3		1565.2			
H3N4F		1606.3			
H4N4F		1647.6			
H5N4F		1929.2	1929.3	1929.1	
H5N5F			2132.0		
H6N5F		2294.7	2294.7		2294.6
H7N6F		2658.6	2658.7		

Table 3: MS analysis of 2AB labelled sCD59 asialo oligosaccharides separated according to charge *(see figs 2a and 8). Numbers are m/z values from figure 8. N(O) / N(C): neutral oligomannose or complex glycans, NA1: N-linked glycans with one charged residue, NA2: N-linked glycans with 2 charged residues. NA2b - contained a biantennary structure, NA2t- contained a triantennary glycan or polylactosamine structure. Additional peaks may represent the underivitised glycans (e.g. peak 1808.9 in 8b is underivitised H5N4F)*

3.7 Molecular modelling of soluble CD59

The relative sizes of the sugars and the protein were explored by molecular modelling (Fig.9). The glycan attached to Asn 18 is the disialylated biantennary complex glycan SA2H5N4F. The length of the glycan is of the same order as the diameter of the protein (approximately 4.2nM)

4. Conclusions

sCD59 contains at least eleven different oligosaccharides of which approximately 94% are complex bi- or triantennary structures with or without polylactosamine extensions and only 6% of unprocessed oligomannose sugars were detected. This suggests that the CHO cell line has efficient glycosylation machinery and that the local protein conformation allows the interaction of GlcNAc transferases I,II and III with the developing glycan chain. The addition of terminal sialic acid prevents the development of lactosamine extensions in approximately 10% of disialylated biantennary structures (Fig.2a). However 12% of the glycans are elongated suggesting that in some cases GlcNAc transferase competes effectively with sialyl transferase. All the complex sugars were present as both neutral (35%) and sialylated (35%:A1; 30% A2) structures although not every galactose residue was sialylated and tri-antennary glycans were

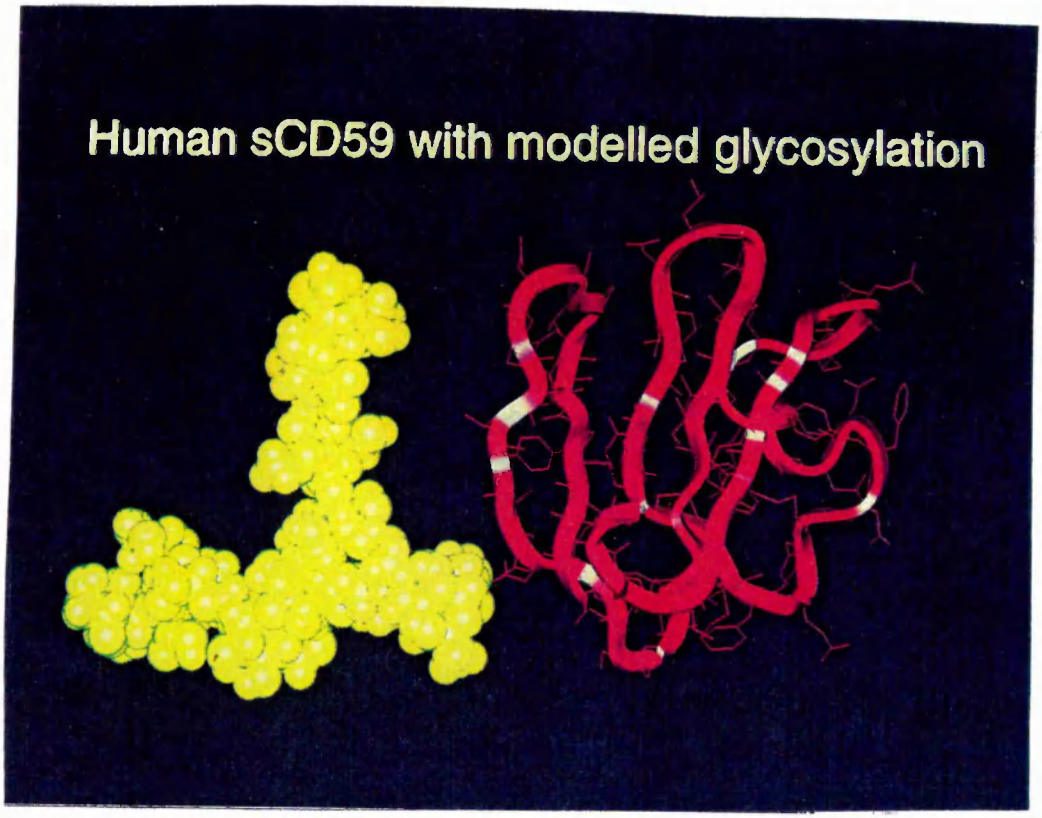


Figure 9: A molecular model of soluble CD59

The protein structure (red) is based on the NMR solution structure (Kieffer 1994). The oligosaccharide (yellow) attached at Asn 18 is the di-sialylated form of the biantennary complex glycan H5N4F which constitutes 41% of the glycans associated with the molecule.

only mono or di-sialylated. Di-sialylated biantennary structures accounted for 12% of the A2 structures while 18% of the A2 glycans were di-sialylated triantennary glycans. The CHO cell line does not express an α-2,6 sialyl transferase, and, consistent with this all the terminal sialic acid linkages associated with soluble CD59 were of the α2,3 type. The incomplete sialylation of polylactosamine structures suggests that elongation of the oligosaccharide as well as the protein structure has a role in controlling the addition of sialic acid to complex glycans. This is consistent with a study of sialylation in a mouse lymphoma cell line, thyroid glycoprotein GP1, and fetal lactosaminoglycan (Merkle and Cummings 1987) which indicates that sialylation is more likely to occur in the early stages of processing and the incidence falls off as the chains become more elongated.

These data indicate that the developing oligosaccharide chain at Asn18 on sCD59, the smallest of the domains of the proteins analysed in this chapter, is accessible to all the glycosylating enzymes. A comparison of the glycosylation of sCD59 with that of other leucocyte antigens is made in the discussion at the end of chapter 2.

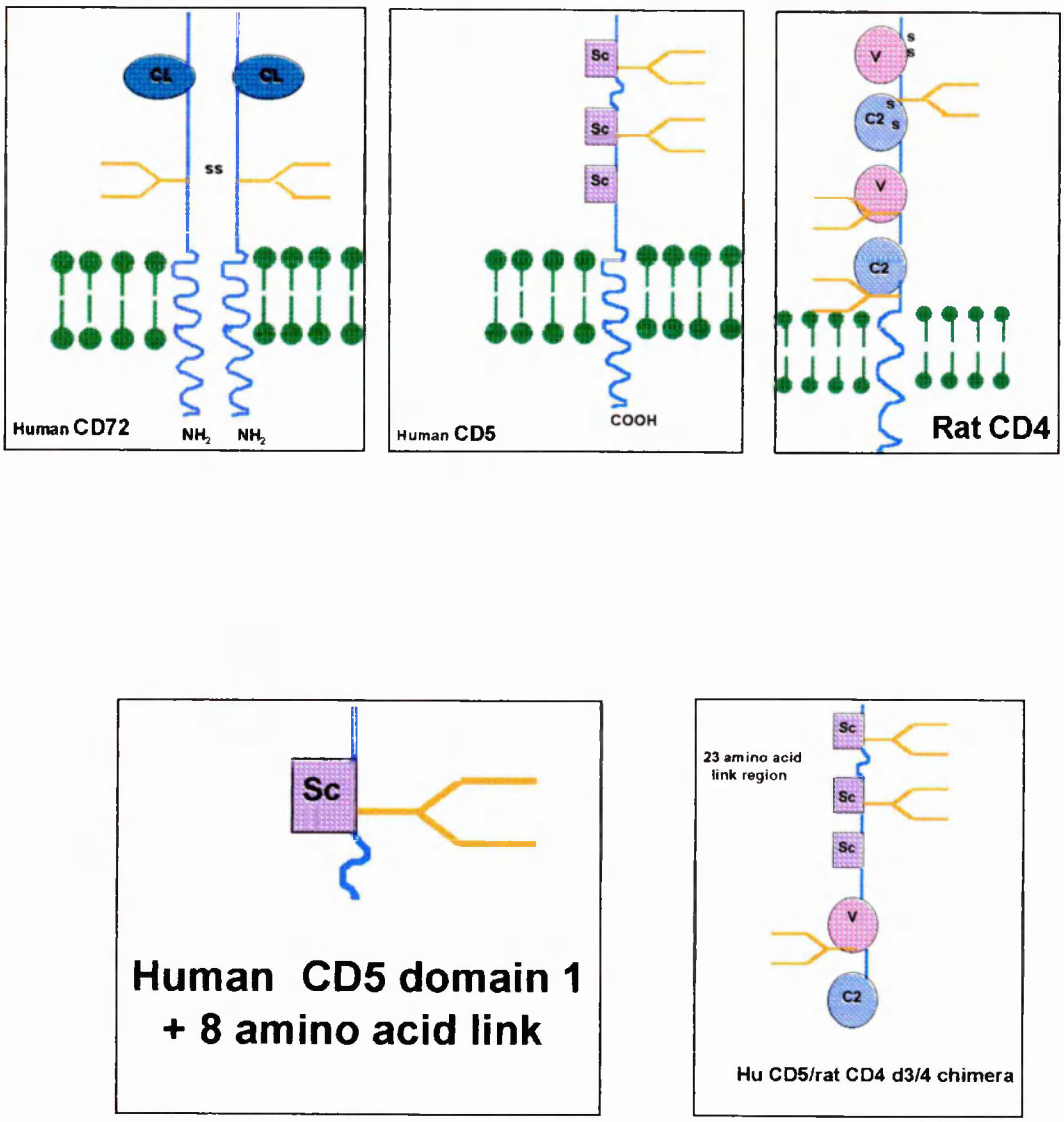


Figure 1: A schematic representation of the molecules and constructs involved in this study.

Ch. 2B The glycosylation of domain 1 of human CD5 and of a chimera of human CD5/rat CD4d3&4, both expressed in Chinese Hamster Ovary cells

1. Background

CD5 (T1, Leu-1, Ly-1) (Fig. 1) (67kD, 471 amino acids) is expressed on all mature T-cells, most thymocytes and on a subset of mature B-cells. The extra-cellular region consists of three scavenger receptor domains (amino acids 1-112, 147- 243, 244-347). The function of scavenger receptor domains is unclear, but they may have a role in macrophage associated immune responses and inflammation (Freeman et al 1990). In general they provide a stable core structure which can withstand the extracellular environment. In CD5 domain 1 is separated from domain 2 by a connecting peptide, or link region, (113-146) which contains 17 threonine and proline residues within a stretch of 20 amino acids (114-133). There are 5 potential N-glycosylation sites of which only the two N-terminal sites, Asn92CysSer and Asn217SerSer in domains 1 and 2 respectively, are occupied. The role of CD5 in immune interactions is still unknown, but it may function as a receptor delivering co-stimulatory signals to T-cells in a manner similar to CD2. Anti-CD5 antibodies stimulate T-cell proliferation (Review: Beyers et al 1989), the secretion of interleukin-2 and its receptor and also induce a rise in intracellular Ca^{2+} concentration. CD5 and CD2 bind to the T cell receptor complex (Beyers et al 1992) to form a loosely associated complex which also involves CD4 or CD8. It has been proposed (Van de Velde et al 1991) that B and T cells communicate through the interaction of CD5 with CD72 (Fig.1), a cell surface glycoprotein found only on B-cells, however this has not been confirmed (Brown, M. and Barclay, N. personal communication). $CD5^+$ B lymphocytes secrete monoreactive IgMRF, and also polyreactive IgM antibodies with broad specificities. In contrast $CD5^-$ cells secrete mainly IgG monoreactive antibodies. Interestingly, $CD5^+$ B cells are upregulated in Rheumatoid Arthritis (RA) when the cells produce increased levels of IgMRF and IgGRF (see ch.6 for glycosylation analysis of IgGRF). Also in RA $CD5^+$ T-cells express a cell surface lectin which binds the O-linked Galβ1-3GalNAc glycans in the hinge regions of IgG and IgD (Rudd et al 1994).

The glycosylated molecules analysed in this chapter are shown schematically in figure1. The glycosylation of CD5 domain 1 (1 N-linked glycosylation site) was compared with that of another single domain protein, sCD59 (ch 2A) and with CD2 (ch 2B) to investigate the influence of the 3-dimensional structure of the protein on glycosylation. In particular, this study explored the glycosylation machinery in CHO cells operating on a single N-linked site (Asn 92) in domain 1 of CD5 (CD5d1) (M.Wt. 17kD) which included the first 8 amino acids of the link region. The link region, which is rich in proline and threonine residues, was examined for the presence of O-linked glycans. Secondly, the glycosylation of a chimera (75kD) consisting of CD5, a link region of 23 amino acids and CD4 domains 3 and 4 (CD4d3/4) (1 N-linked glycosylation site) was analysed. This explored the consequences for the glycosylation of adding four further domains to CD5d1. Finally, to explore the possibility that the insertion of link regions into

recombinant proteins may be a means of creating O-linked domains in the CHO cell expression system, the extent of O-glycosylation of an 8-amino acid C-terminal region attached to CD5d1 was compared with the O-glycosylation of intact CD5 attached to domains 3 and 4 of CD4 which contains a 23 amino acid link region between CD5d1 and 2.

2. Strategy:

Glycans were released by hydrazinolysis and fluorescently labelled with 2-aminobenzamide (2AB) or sodium amino napthalene trisulphonic acid (ANTS) or radiolabelled with tritium. The relative amounts of neutral glycans (N), N-linked glycans containing a single charged group (NA1) and N-linked glycans with 2 charged groups (NA2) were determined by weak anion exchange (WAX) chromatography and fluorophore assisted carbohydrate electrophoresis (FACE).

The linkages and type of the charged species were characterised by determining the susceptibility to *Arthrobacter Ureafaciens* and Newcastle Disease Virus sialidases. Asialo glycans were resolved by P4 GPC and FACE (FACE data in ch.8 2.4). N-linked glycans isolated from P4 GPC were characterised by RAAM analysed by P4 GPC (data in ch.8 1.8) and by MALDI MS (data in ch.8 2.1.ii). Asialo N-linked glycans were analysed by MALDI MS and structures were determined by exoglycosidase sequencing followed by MALDI MS. Structural analysis of the O-linked glycans was by exoglycosidase sequencing followed by P4 GPC and the end group analysis of the O-linked glycan was carried out by radioactive Gas Chromatography (GC).

The oligosaccharide sequencing methods used here are described in chapter 8 of this thesis.

3. Results

The CD5d1 and CD5/CD4d3,4 constructs were prepared by *in vitro* mutagenesis. The proteins were purified by affinity chromatography and gel filtration (Dr. M. McAllister and Dr. M. Brown).

Analysis of oligosaccharides attached to CD5d1

3.1 Release and labelling of glycans from CD5d1

Sugars (85 nmoles assuming 85% release and recovery from the GlycoPrep machine) were released from CD5d1 (1mg) by hydrazinolysis using the GlycoPrep 1000 (OGS) under conditions optimised for maximum recovery of N- and O-links. The glycans were evaporated to dryness and re-dissolved in 200µl water. 75 nmoles of CD5d1 sugar was labelled at the reducing terminus with tritium (total counts 4.4E7) according to standard procedures. 1nmole was labelled with sodium amino napthalene trisulphonic acid (ANTS) and analysed by Fluorophore Assisted Carbohydrate Electrophoresis (FACE). 10nmoles were labelled with 2-aminobenzamide (2AB); the remainder was unlabelled and used for the Matrix Assisted Laser Desorption Mass Spectrometry (MALDI MS) analyses. 80% of the glycans from the chimera were radiolabelled (total counts 1.7E8), 2% was used for the FACE analysis and 5% was labelled with 2AB. The remainder was unlabelled and used for the mass spectrometric analysis.

3.2 Charge analysis of the glycans released from CD5

3.2.1 WAX chromatography

Figure 2a shows the resolution of five populations of CD5 glycans on a WAX chromatography column separated according to charge (see table 1). On digestion with *Arthrobacter ureafaciens* neuraminidase all of the glycans were reduced to neutral structures (Fig 2b) indicating that all the charge was due to either α1,3 or α1,6 linked sialic acid. The small peak eluting at fraction 35 is an artefact of the radiolabelling procedure which does not elute with any standard glycan. It is not present in figure 2c where the glycans have been fluorescently labelled with 2-aminobenzidine (2AB).

Complete digestion of the charged glycans to neutral structures by Newcastle Disease Virus (NDV) neuraminidase (Figure 2c) specific for α 2,3 (or α 2,8) linkages confirmed that all the charge was due to α1,3 linked sialic acid and indicated that 85% of the sugars recovered in this pool were N-linked. This is in close agreement with the data from figure 2a. Comparison of the data in figures 2a and 2c shows that there are no naturally neutral O-linked sugars associated with CD5.

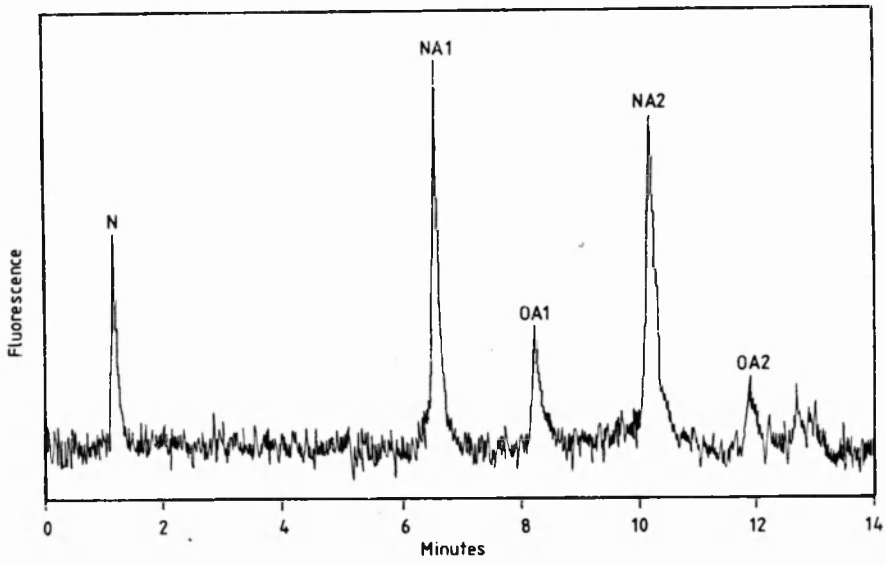


Figure 2a: Analysis of the hydrazine released pool of 2AB labelled CD5 glycans by weak anion exchange chromatography.
Assignment of peaks was by comparison with standard fetuin sugars. N, NA1, NA2 neutral, mono- and di-sialylated N-linked glycans respectively. OA1,OA2: mono- and di-sialylated O-linked glycans, respectively.

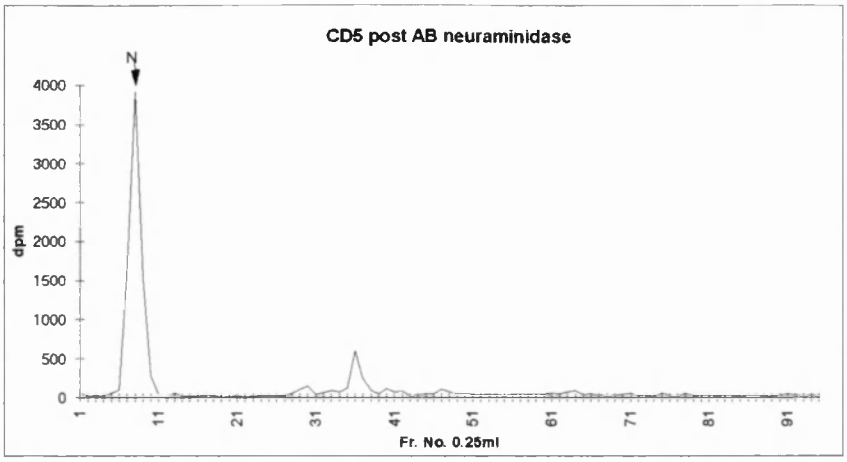


Figure 2b: Weak anion exchange chromatography analysis of tritium labelled CD5 glycans after digestion with *Arthrobacter ureafaciens* neuraminidase

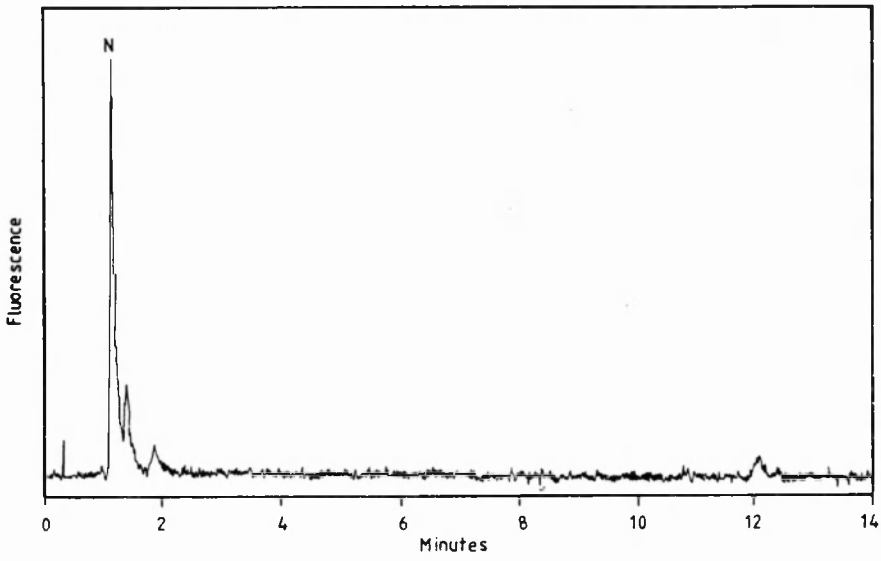


Figure 2c: Weak anion exchange chromatography analysis of 2AB labelled CD5 glycans after digestion with Newcastle Disease Virus neuraminidase

The analysis of the tritium labelled sample in figure 2b shows a radiolabelled contaminant which is not present in the 2AB labelled sample analysed in figure 2c. This contaminant may account for the discrepancy in the percentages of N:O-links calculated from P4 (radiolabelled) and WAX (2AB labelled) data in Table 2.

The charged sugars were also analysed by Fluorophore Assisted Carbohydrate Electrophoresis (FACE) to compare the results of a gel electrophoresis separation with the separation by HPLC. The data (ch 8 section 2.4) confirm that two populations of charged sugars can be reduced to a single species by incubating with *Arthrobacter ureafaciens* neuraminidase.

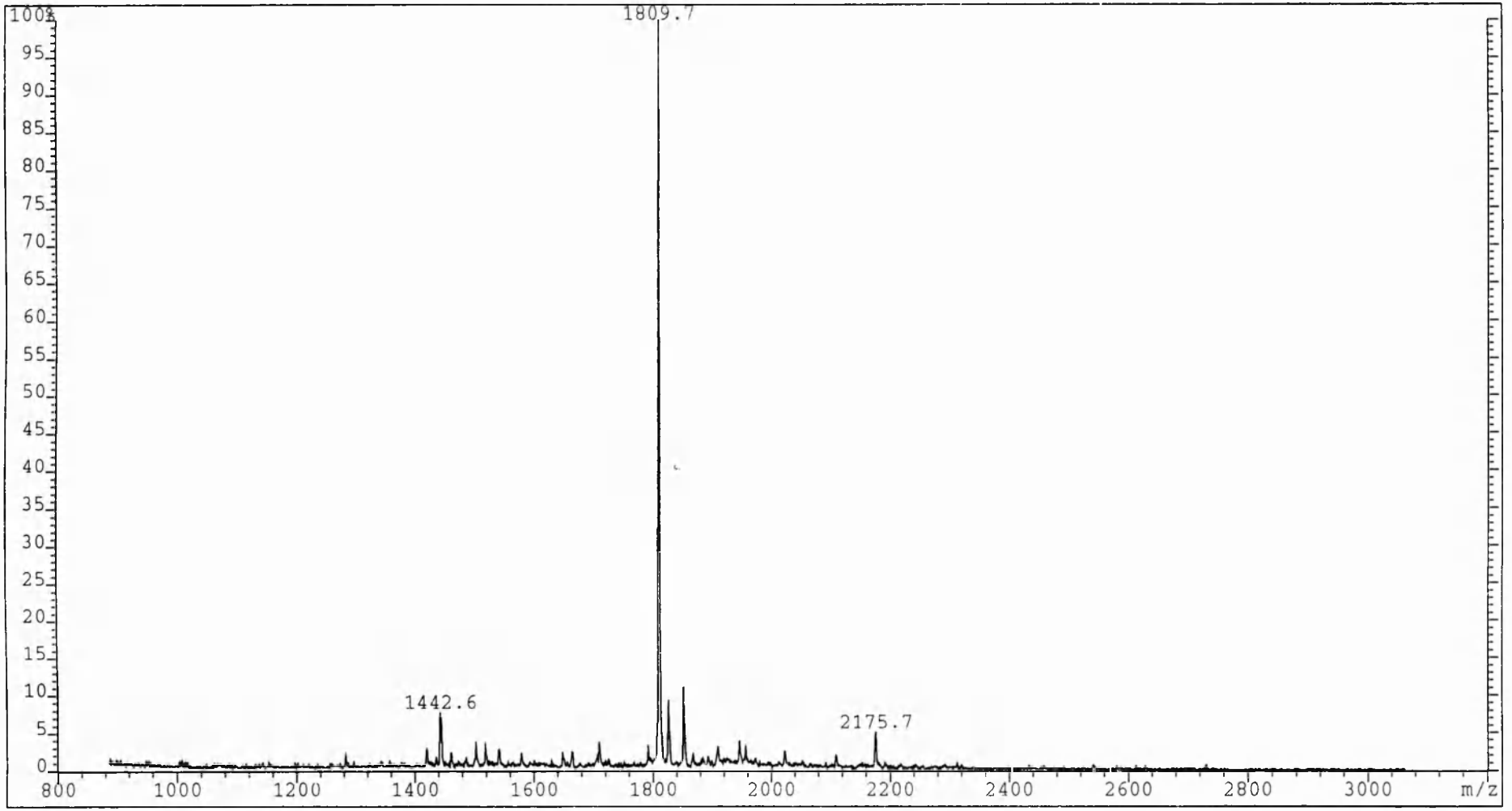


Figure 5a *CD5 asialo N-linked glycans analysed by MALDI MS*
Figure 5a shows the MALDI MS analysis of the asialo glycan pool from figure 1b over the range

N N-links	N	O-links	NA1	NA2	OA1	OA2
(i) **WAX (2a)**	--------15%-------		32%	36%	12%	5%
(ii) **WAX (2c)**	85%	15%				
(iii) **P4 (3)**	66%	33%				

Table 1: A comparison of the charge profiles of CD5 obtained by different methods. *(i) The percentages of the charged (NA1, NA2, OA1, OA2) and neutral (N N or N O) populations were calculated from figure 2a. The percentages of neutral N- and O- links cannot be separated. (ii) The ratio of total N : O-links taken from the WAX analysis (fig. 2b). Percentages were obtained by integration of the WAX data. (iii) Ratio of N:O-links calculated from integration of the P4 data in figure 3.*

3.3 Size analysis of asialo glycans from CD5

3.3.1 Analysis of CD5 asialo glycans by P4 GPC

The asialo glycans from figure 2b were resolved by P4 GPC (Fig.3). Two peaks were obtained, at 14.8 and 3.5 glucose units in the ratio 66:33 respectively.

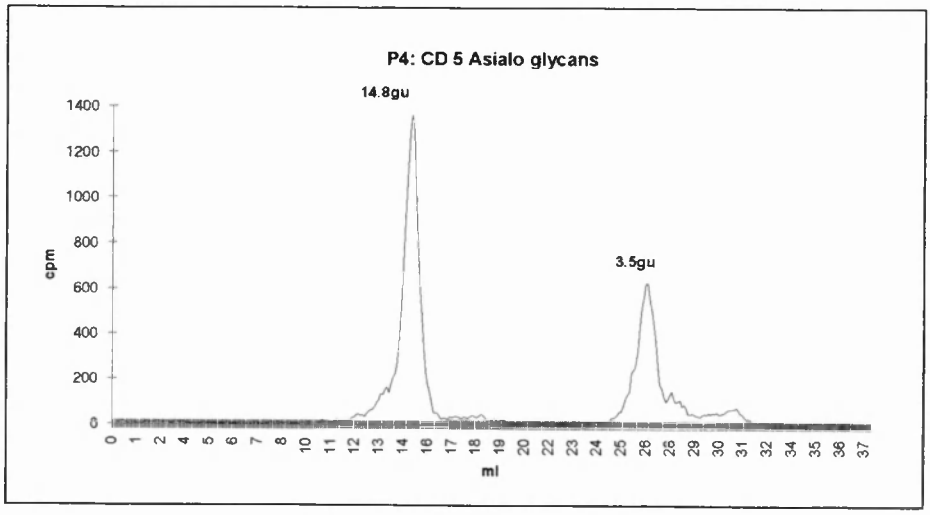


Figure 3: P4 GPC of asialo CD5d1 glycans

3.3.2 Analysis of the CD5 neutral glycan at 14.5gu by RAAM.

The pool of glycans eluting at 14.8gu (Fig.3) was analysed using the standard RAAM N-glycan enzyme array. The products of the digestion were analysed by P4 GPC and by MALDI MS (see ch.8 section 1.8 and ch. 8 section 2.1(ii) respectively for the data). The glycan was identified as the core fucosylated biantennary complex glycan, with the structural formula H5N4F, shown in figure 4.

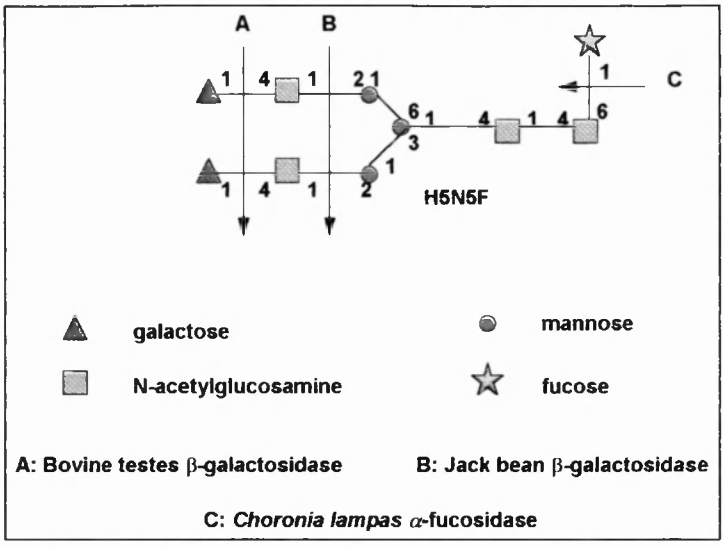


Figure 4: The N-linked 14.8gu glycan attached to CD5 identified by RAAM.

The arrows show the cleavage positions of the enzymes used in the sequential exoglycosidase digestions. The products were analysed by MALDI MS (see figure 5a-d) to confirm the structure.

3.3.3 Structure of the CD5 asialo N-linked glycans confirmed by MALDI MS and sequential exoglycosidase digestions

Figure 5a shows the MALDI MS analysis of the asialo glycan pool from figure 1b over the range 900-3500 mass units. The peak at 1809.7 mass units is consistent with the H5N4F structure identified by RAAM and shown in figure 4. Figures 5b,c and d are the analyses of the products of digestions A,B, and C 3 respectively. The peaks are at 1485.6, 1079.1 and 933.1 mass units, consistent with the structures H3N4F, H3N2F and H3N2.

3.4.1 Analysis of the O-linked glycan eluting at 3.5gu in figure 2 by exoglycosidase digestion

The 3.5gu (figure 3) structure was digested with Bovine Testes β-galactosidase and the product was analysed by P4 GPC (Fig.6). The glycan was reduced by 1gu to a structure eluting at 2.5gu consistent with the presence of an O-linked glycan Galβ1-3/4-GalNAc.

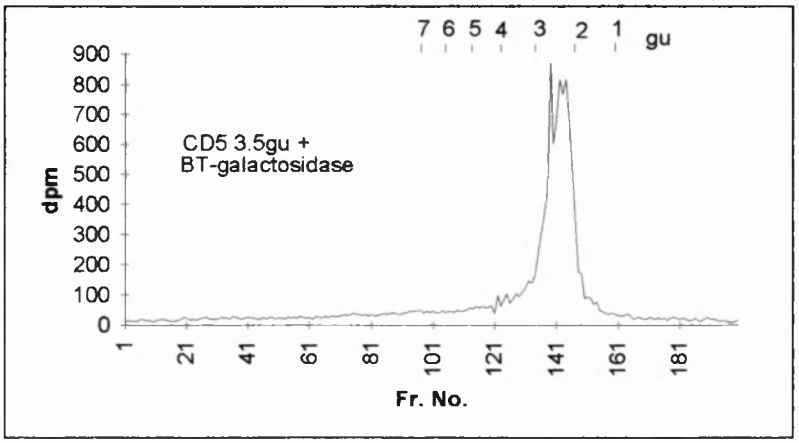


Figure 6. P4 GPC profile of the product of the digestion of the CD5 O-linked glycan Galβ1-3/4GalNAc with Bovine Testes galactosidase

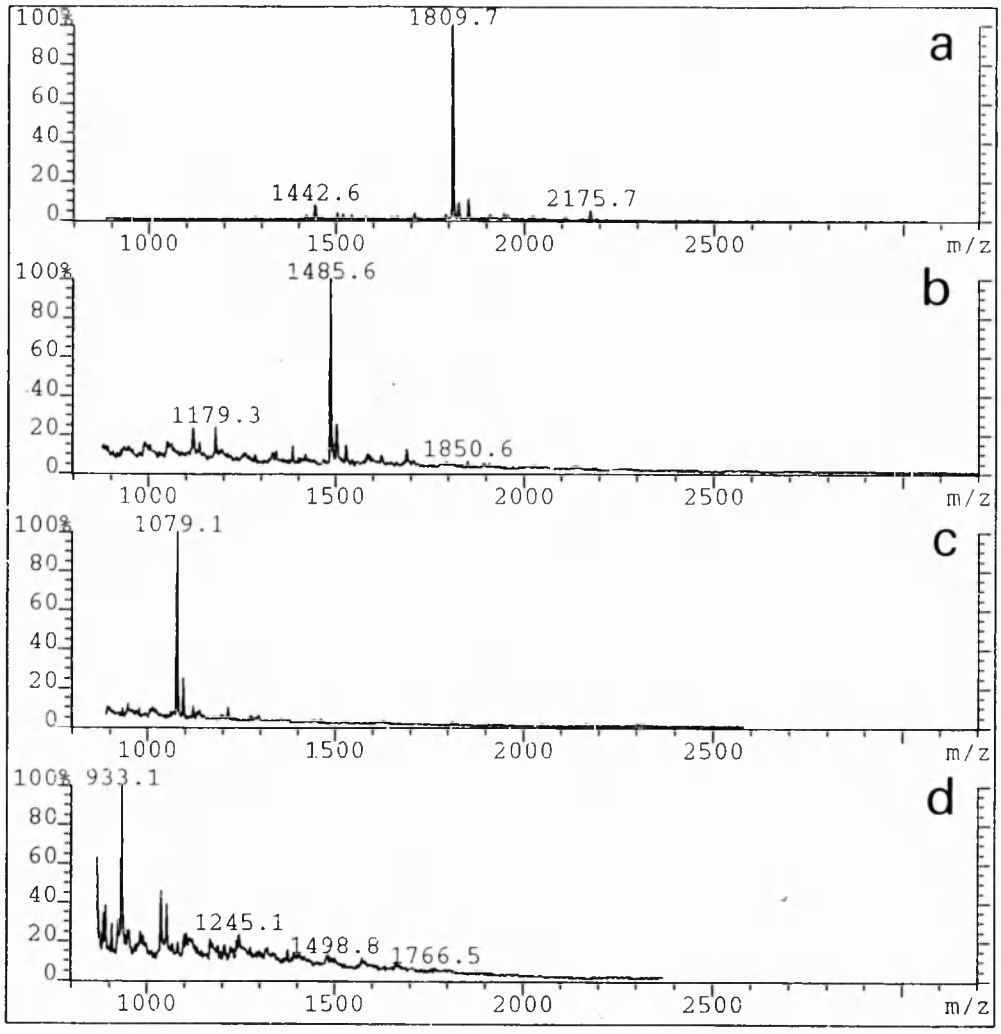


Figure 5a-d: 5a is the MALDI MS analysis of the glycan pool (H5N4F). 5b,c and d are the analyses of the products of digestions A,B and C respectively (see figure 4). The peaks at 1485, 1079 and 933 mass units are consistent with the structures H3N4F, H3N2F and H3N2 respectively.

3.4.2 Analysis of the sugar at the reducing terminus.

The reducing terminal glycan was identified by gas chromatography (GC) equipped with a radioactive detection system. The 2.5 gu peak from figure 6 was co-injected into the GC column with radiolabelled inositol, used as an internal marker (Fig. 7). In this technique structures are assigned by comparing the difference in the elution times between the unknown sample and inositol with the differences between standard sugars and inositol. The 2.5 gu peak eluted at the same position as N-acetyl galactosamine (GalNAc), consistent with the presence of an O-linked glycan terminating in GalNAc.

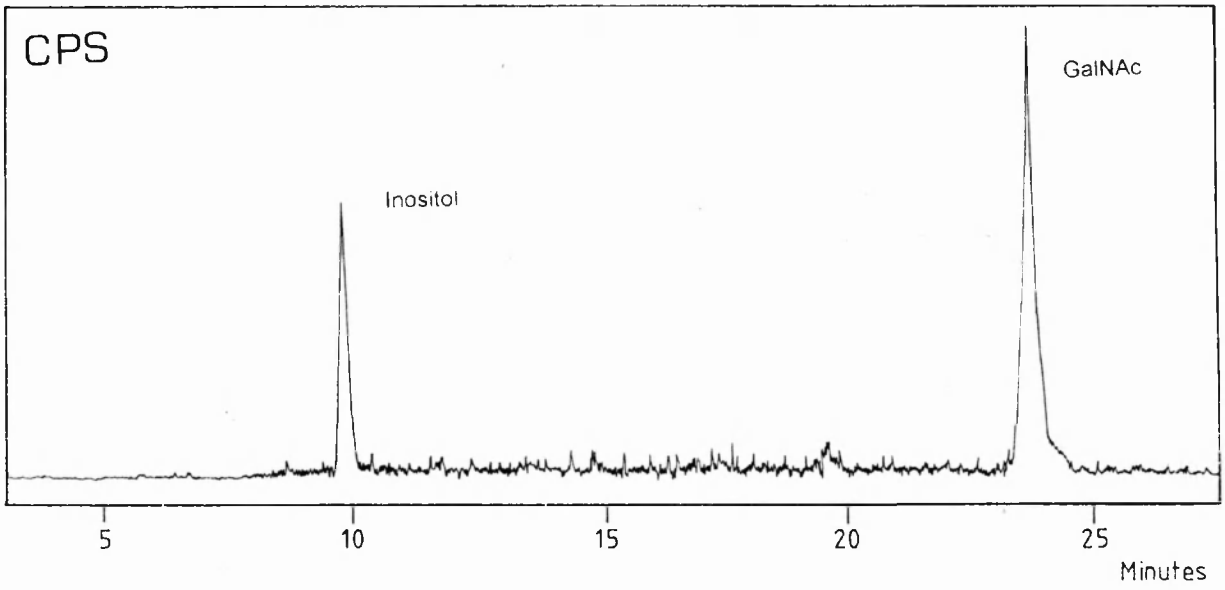


Figure 7: The radioactive GC analysis of the terminal glycan on the CD5 3.5gu glycan

4. Results of analysis of glycans attached to the CD5/CD4d3,4 chimera:

Glycans were released by hydrazine at 95^{o}C and labelled with 2AB or tritium. The relative amounts of N, NA1 and NA2 were determined by WAX chromatography and the charged species characterised by digestion with *Arthrobacter Ureafaciens* sialidase. Asialo glycans were resolved by P4 GPC and MALDI MS and analysed by RAAM followed by P4 GPC and MALDI MS. O-linked glycans were analysed by exoglycosidase sequencing and P4 GPC and end group analysis of O-linked glycan used radioactive GC.

4.1 Release and labelling of glycans

The glycans (10.2nmoles, estimated 85% yield) were released from 300μg of the CD5/CD4d3,4 chimera (M.Wt. 75kD). 18% was used for MS studies, 2% was labelled with 2AB and the remainder (9.6nmoles) was radiolabelled with tritium (1.7E8 cpm in total).

4.2 Analysis of CD5/CD4d3/4 by charge

The WAX chromatography analysis of 2AB labelled glycans (Figure 8a) shows that none of the glycans are neutral, 25% are monosialylated N-linked glycans, 46% are disialylated N-links and

29% are monosialylated O-links and 0% are disialylated O-links (compared with 15%, 32%, 36%, 12% and 5% for CD5d1 alone). The ratio of N : O-links was 71.4% : 28.5% (compared with 85% : 15% in CD5d1). All structures were reduced to neutral with *Arthrobacter ureafaciens* sialidase (Fig. 8b) indicating that all the charge is due to sialic acid.

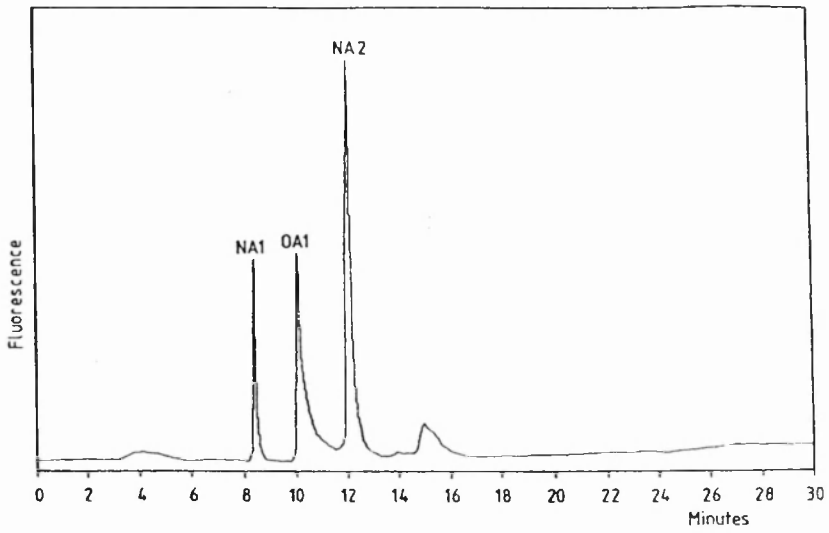


Figure 8a: WAX chromatography analysis of 2AB labelled glycans released from CD5/CD4d3/4

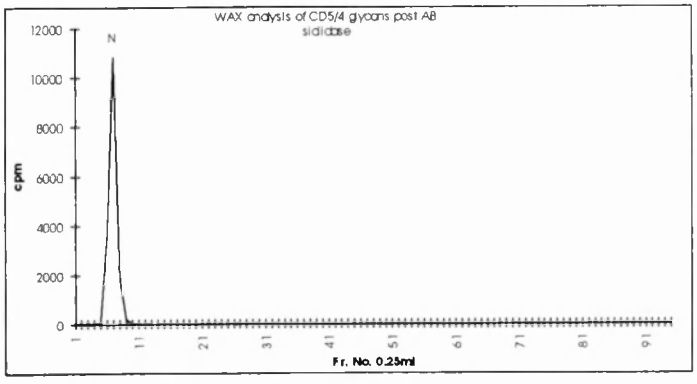


Figure 8b: WAX chromatography analysis of asialo tritium labelled glycans from CD5/CD4d3/4

4.3 Analysis of CD5/CD4d3/4 asialo N- and O- glycans by P4 GPC and radio GC

Figure 9 shows the P4 profile of the asialo glycans released from CD5/CD4d3/4. 92% of the N-linked glycans eluted at 14.6gu and were identified by RAAM and MALDI MS as the core fucosylated biantennary complex sugar, H5N4F, found on CD5. The 3.6gu O-linked sugar was also identical to the glycan on CD5d1. It was cleaved with β-galactosidase to a 2.5gu structure which was shown by radioactive GC to be N-acetyl galactosamine. The material eluting at 0.5 gu was not analysed and may be an environmental contaminant since the molecular volume is less than that of a monosaccharide. There was a small amount of material eluting at more than 14.6gu, but insufficient to be analysed.

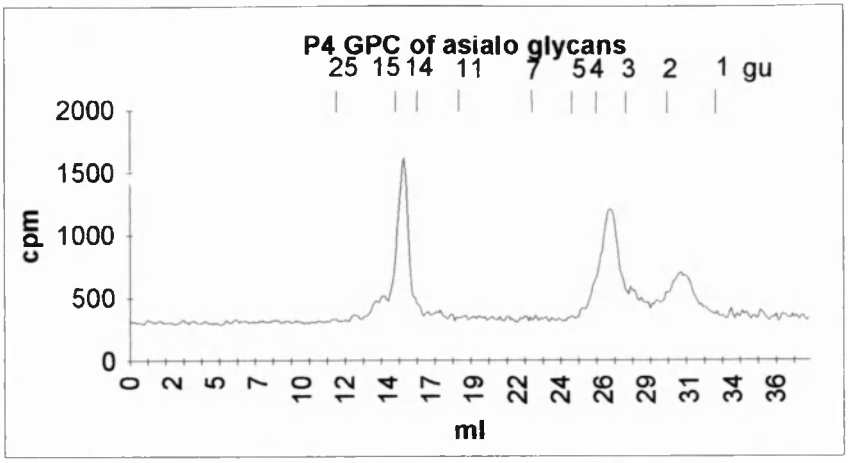
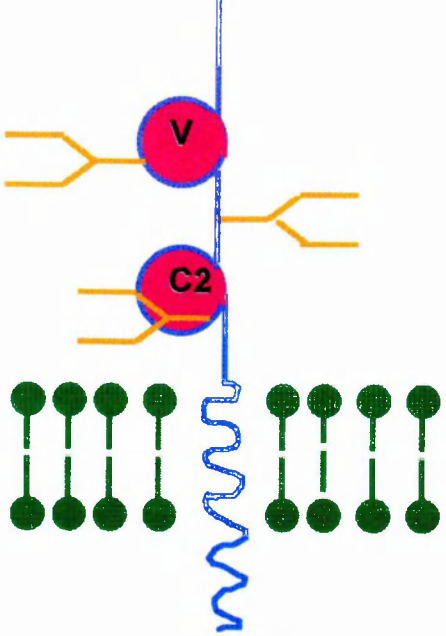


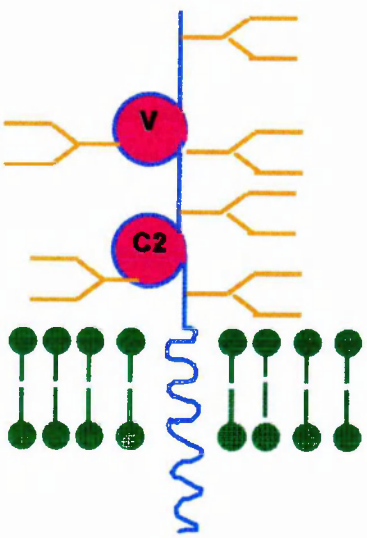
Figure 9: P4 GPC of the asialo glycans attached to CD5/CD4d3/4. The peaks are at 14.6, 3.6 and 2.5gu

5. Conclusions

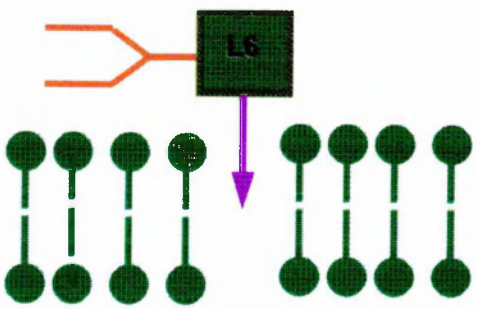
A comparison of the P4 profiles for the two molecules released under the same conditions shows that CD5/CD4d3/4 has a higher proportion of O-linked sugars than CD5d1 (Figs. 3 and 9). This may reflect the increased length of the link region and the higher proportion of Thr and Pro residues in the chimera. Interestingly, there have been no previous reports of O-linked sugars attached to CD5. In neither case is there evidence of significant quantities of N-glycans other than the core fucosylated bi-antennary complex glycan, H3N4F, commonly found on serum glycoproteins. This homogeneity of glycosylation is extremely unusual and in direct contrast to the diversity of structures found on sCD59, which is also a single domain protein. This suggests that there is a difference in the control exerted over the glycosylation pathway by the scavenger domain in CD5d1 and the Ly6 domain in CD59. The fact that the same single glycan is also present on all three sites of the chimera, (which is fully glycosylated - Brown, M. and Barclay, N. - personal communication) is also remarkable. It implies that all sites are equally accessible to the enzymes, and that there has been no site specific processing although both scavenger and Ig-like domains are involved. The data suggest that the additional domains do not shield the glycans in CD5d1 from the processing enzymes and this is consistent with the suggestion that the five domains are linear (Dr. N. Barclay - personal communication). A comparison of CD5d1 with the chimera shows an increase in sialic acid content in the chimera. The CD5/CD4d3,4 chimera contains no neutral structures while 17% of the glycans in CD5d1 are neutral bi-antennary complex sugars. Interestingly, a study of the glycosylation of the intact rat CD4 protein (Ashford et al 1993) identified neutral glycans. Moreover, in contrast to the CD4 chimera, no core fucosylation of rat CD4 was observed. This is consistent with the well established finding that glycosylation is cell type specific and that the protein may control its own glycosylation. The quaternary structure of the protein may also be important, but this has not been established. In the following section the glycosylation of CD2 was analysed. This protein consist of two domains, both of the IgSF class.



Human CD2



CD58



Human erythrocyte CD59

Figure 1a: CD2, its ligand CD58 and possible ligand CD59.

Ch. 2C The glycosylation of human CD2 and rat CD48

1. Background

CD2 (T11, LFA-2) is expressed on most T-lymphocytes, NK cells and thymocytes. It is a membrane bound protein with a transmembrane anchor (Fig.1a) and plays a role in mediating adhesion between T-cells and target cells (Qin et al 1994). In humans proposed ligands include the cell surface antigens CD58, CD48 and CD59 while in rat CD2 binds CD48 (van der Merwe et al 1995). CD2 interacts with the T-cell receptor complex (TcR/CD3). The structures of both rat and human soluble CD2 have been determined by X-ray crystallography (Jones et al 1992, Bodian et al 1994), when the molecules formed dimers through head to head interactions between the amino-terminal domains. CD2 and its proposed ligands CD58 and CD48 are part of a subgroup within the immunoglobulin superfamily (IgSF) and their genes are linked in the genome indicating that they may have evolved by gene duplication. The extracellular domain of CD2 (M.Wt. 36.8kD) (Fig.1b) is made up of two IgSF domains each of which contains one N-glycosylation sequon at N65GT and N126GT. The two domains are joined by a linker region which also has a single N-glycosylation site at N117TT. The amino acid sequence contains many serine and threonine residues, however no O-linked sugar has been reported.

(i) Interaction of CD2 with CD58 (LFA3)

CD2 interacts through its amino terminal domain with CD58 to form an adhesion pair, for example between T-cells and cytotoxic targets or antigen presenting cells (Selvaraj et al 1987). Such recognition events have a role in the process of cell-mediated cytotoxicity, antigen and mitogen-induced T-cell proliferation and lymphokine production. Through a series of mutations it has been shown that the binding site for CD58 lies on one face of the first domain of CD2 (Somoza et al 1992). This active site is on the opposite side of the molecule to the glycans at Asn 65.

(ii) The role of glycosylation at Asn 65 in the CD2/CD58 interaction.

In studies of CD2 deglycosylated with PNGase F, N-glycosylation was shown to be required for human CD2 immunoadhesion functions (Recny et al 1992). Neither deglycosylated CD2, nor the mutant CD2 molecule containing a single Asn65 to Gln65 substitution, bound CD58 or monoclonal antibodies directed at native CD2 adhesion domain epitopes. However, rat CD2 containing only a single GlcNAc residue at each of the glycosylation sites showed no decrease in activity or antigen binding (van der Merwe et al 1993).

(iii) Interaction of CD2 with CD48

CD48 is reported to be a ligand for CD2 in murine and rat cell systems (Sandrin et al 1993, van der Merwe et al 1993). CD48 (M_r 48kD) is present on leucocytes and, like CD58, has a GPI anchor. In many other respects it is similar to CD58: for example, it has 68% similarity at the amino acid sequence level, and similar polypeptide length, hydrophobicity and predicted secondary structure. It has a low affinity for CD2 (IC_{50} : 10^{-3}M determined by

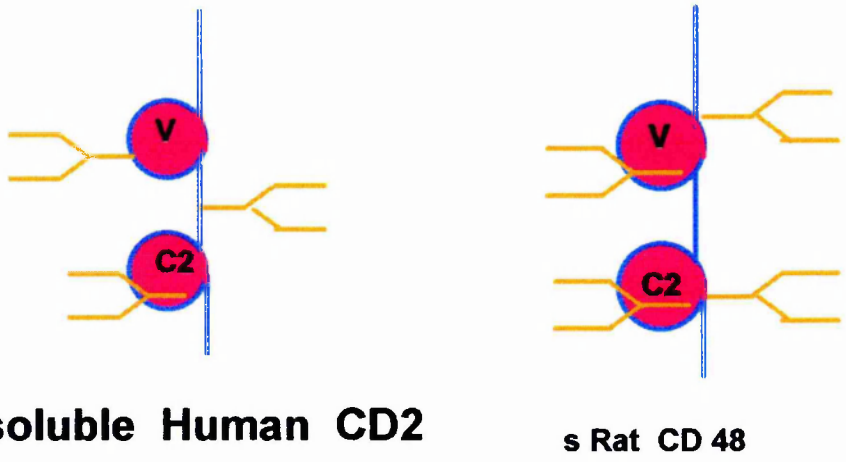


Figure 1b: Soluble CD2 and CD48 analysed in this study.

The molecules were expressed without the transmembrane section with the result that they were secreted into the tissue culture medium.

inhibiting the binding of peripheral blood lymphocytes to CD48 transfectants by either CD48 mAb pre-treatment of COS cells or CD2 mAb pre-treatment of preipheral blood lymphocytes) (Sandrin et al 1993) compared with CD58 (IC 50:$<10^{-5}$ M) (Parish et al 1993). CD58 and CD48 interact with the $T11_1$ epitope of CD2 suggesting that CD48 and CD58 are alternative ligands for the same region of CD2.

(iv) Interaction of CD2 with CD59.

CHO cells stably transfected with human CD59 form rosettes with human T-cells and these can be inhibited by a CD59 monoclonal antibody (Deckert et al 1992). However, the interpretation of this result is unclear, and binding studies using plasmon surface resonance have been unable to demonstrate this interaction at the molecular level (van der Merwe et al 1994).

(v) Interaction of CD2 with the T-cell receptor complex.

Normal T-cells express a multimolecular complex consisting of CD4 or CD8, TcR/CD3, CD2 and CD5 in association with internal tyrosine kinases. It has been proposed that the interaction of CD2 with CD58 (LFA3) may regulate transmembrane signalling by interacting with this multimolecular complex (Beyers et al 1992).

2. Strategy

In this chapter the glycans attached to human soluble CD2 (Fig. 1b) were released by hydrazinolysis, labelled with 2AB or tritium and analysed on the basis of charge by WAX chromatography. Asialo glycans were analysed by a combination of P4 GPC, MALDI MS and exoglycosidase sequencing. Glycans released by hydrazinolysis from rat CD48 were desialylated and analysed by MALDI MS. Molecular models of CD2 and CD48 were constructed.

3. Results

3.1 Preparation of CD2 and release of glycans

The CD2 which was used in this study was expressed without the transmembrane and polar regions of the peptide and was secreted into the medium as a soluble protein.

2mg of human CD2 expressed in CHO cells and purified by affinity chromatography and gel filtration (Dr. Simon Davis, William Dunn School) was dialysed into 0.1%TFA. Glycans (68nmoles - assuming 85% release and recovery as determined by OGS) were released by hydrazinolysis at 95^{o}C under conditions designed to cleave N- and O- glycosidic bonds and re-N-acetylated using the GlycoPrep 1000 (OGS). 80% of the released sugars were labelled with tritium, 5% was used for mass spectrometric analysis, 1% for FACE analysis and for 5% for preliminary studies involving the analysis of fluorescently derivitised sugars. 9% was retained unlabelled.

3.2 Analysis of CD2 glycans by charge using WAX chromatography

Fig. 2a is the charge profile of the tritium labelled glycans released from CD2 and analysed by weak anion exchange (WAX) chromatography. On treatment with *Arthrobacter ureafaciens* neuraminidase all the structures become neutral (Fig. 2b) indicating that all the charge is due to sialic acid. Figure 2c is the analysis of the same pool of released glycans labelled with 2-aminobenzamide (2AB) and with on line fluorescence detection. A comparison of figures 2a and 2c shows that the 2AB labelling procedure is highly specific for sugars compared with sodium borotritide reduction. This procedure may result in the radiolabelling of peptide fragments remaining after hydrazinolysis and it is common for artefacts to be detected as in figure 2a. The glycan peaks were assigned by reference to standard fetuin sugars.

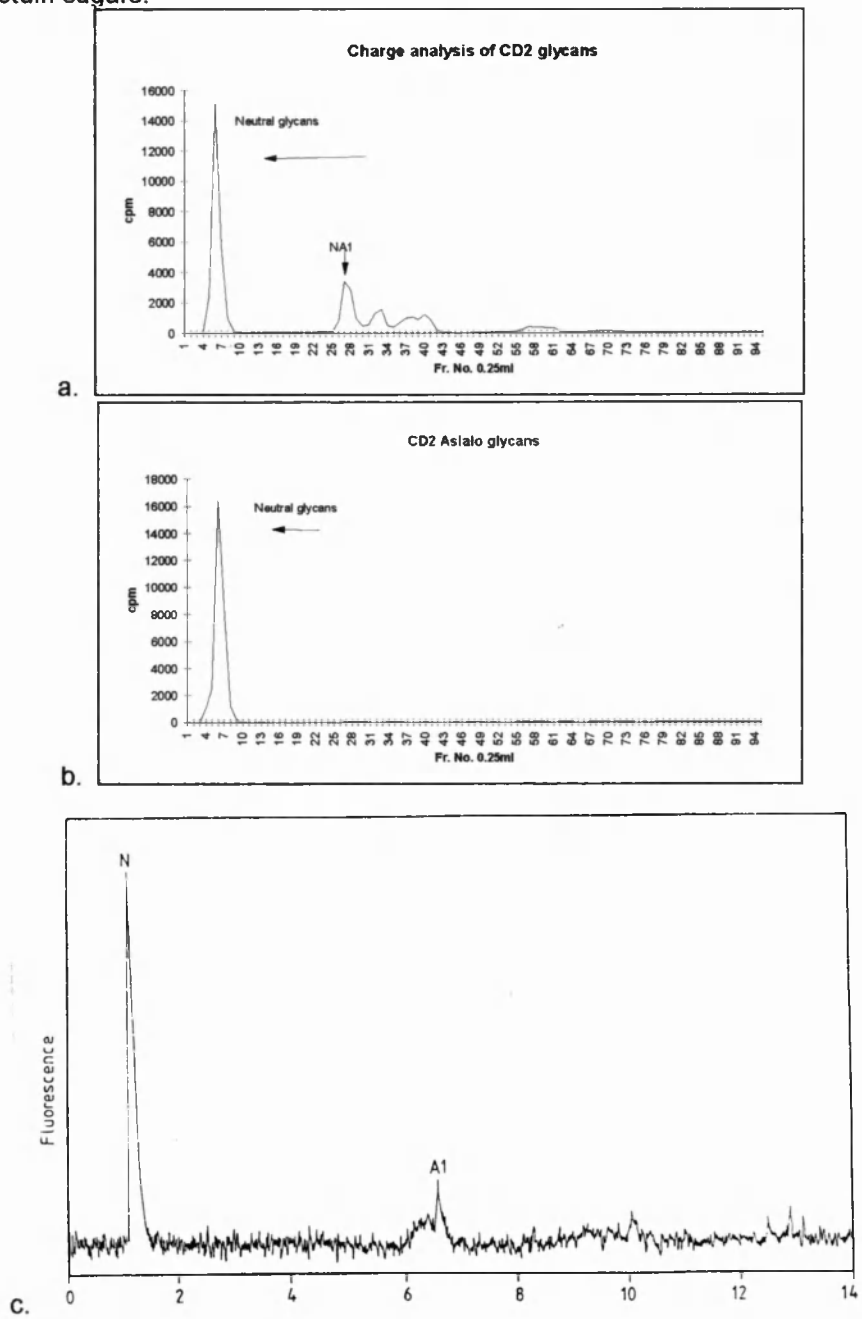


Figure 2a-c: Analysis of the radiolabelled CD2 glycan pool before (a) and after (b,c) digestion with *Arthrobacter ureafaciens* neuraminidase to remove sialic acid. *The*

sugars were resolved by WAX chromatography (Vydac 301VHP575) and the peaks assigned by comparison with fetuin sugars resolved on the same column. 90% of the CD2 glycans are neutral and all the charge is due to sialic acid. Fig.2c shows the separation of the same charged glycan pool under the same conditions as in fig.2a. In 2c the oligosaccharides were labelled with 2AB.

Analysis of CD2 glycans by charge
N:90% NA1: 10%

3.3 Analysis of asialo CD2 glycans using a combination of P4 Gel Permeation Chromatography (GPC) and exoglycosidase sequencing.

3.3.1 Resolution of asialo CD2 glycans by P4 Gel Permeation Chromatography

Figure 3a is the P4 profile of the asialo CD2 glycan library derived from the released glycans after treatment with neuraminidase (Fig. 2b). Eleven peaks were resolved between 25 and 8 glucose units. No glycans in the size range which is exclusive for O-linked structures (between 2.5 and 8.9 gu) were detected. The data are summarised in Table 1.

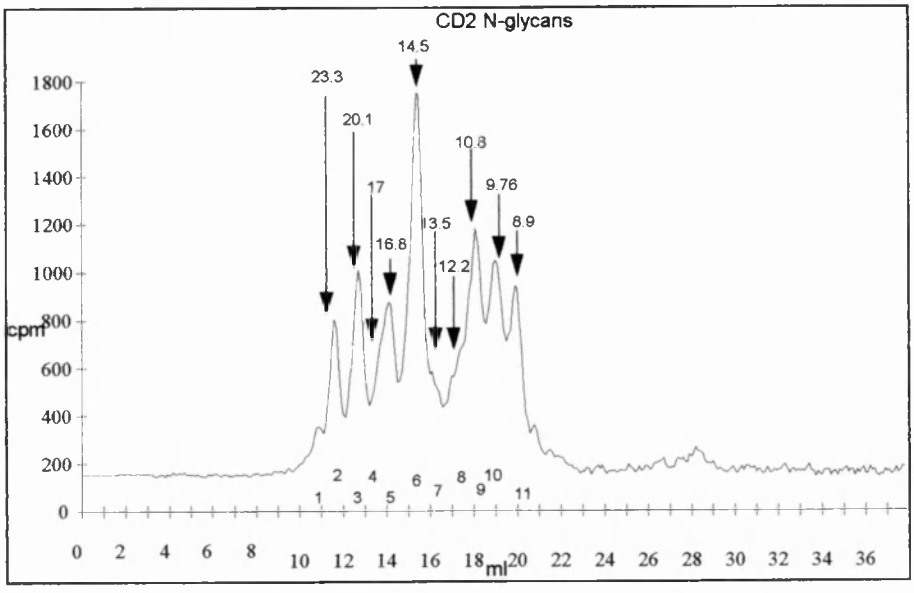


Figure 3a: P4 GPC separation of the neutral glycans in fig. 2b

The sizes of the peaks are labelled in glucose units and the peaks are numbered along the x-axis from 1-11.

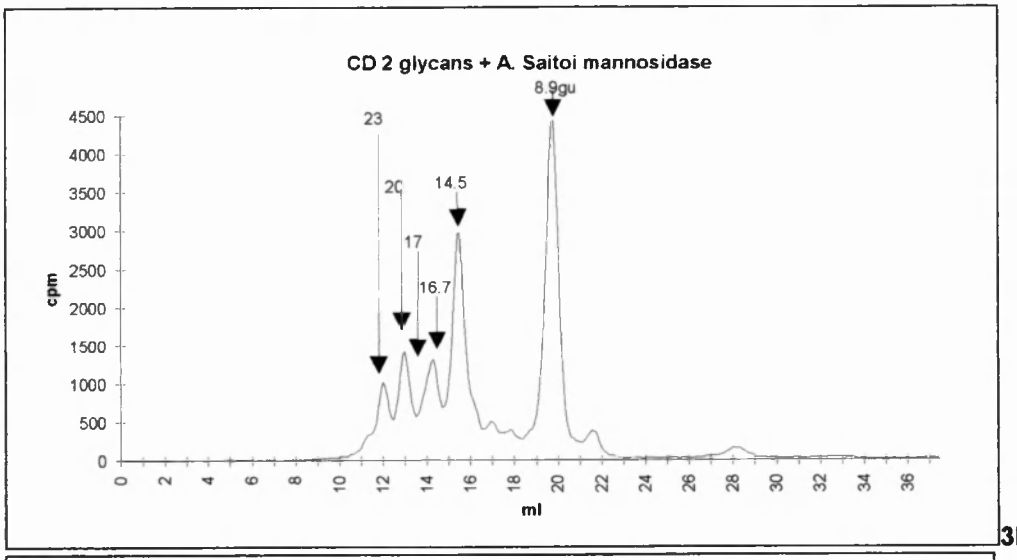
3.3.2 Exoglycosidase analysis of CD2 oligomannose structures

Fractions of the pool of desialylated N-linked glycans were incubated with (i) *A. Saitoi* α-1,2 mannosidase and (ii) Jack Bean α-mannosidase which cleaves terminal, non-substituted, α-D-1,2,6>3 mannose residues. The digests were analysed by P4 gel filtration and the profiles compared with the original analysis of the glycan pool (Fig. 3a).

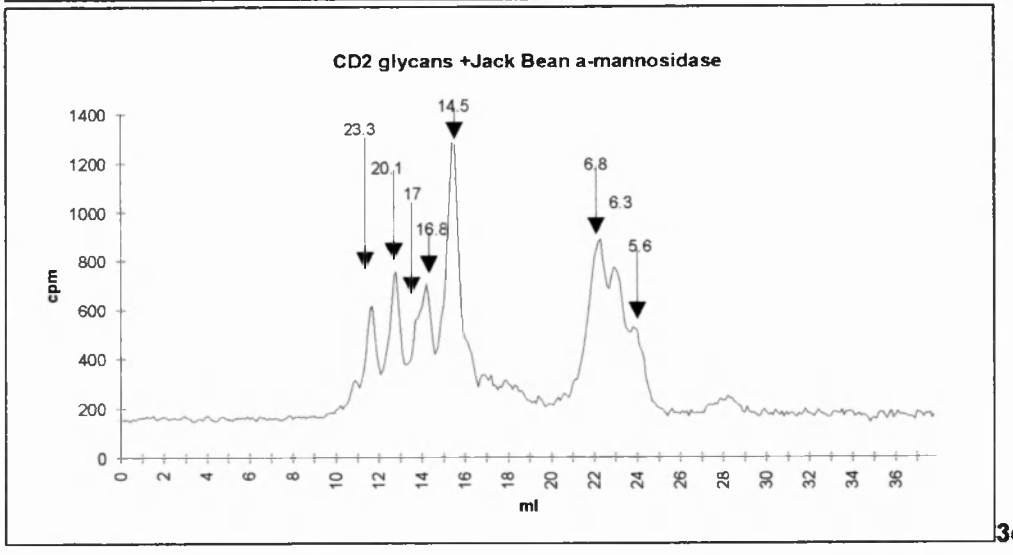
Glycan	Ch. mass	G.U.	%
M5	1257.9	8.85	10
M6	1418.9	9.79	12
M7	1581.6	10.82	18
M8	1743.74	12.16	3
M9	1905.88	13.84	3
H5N4F	1809.7	14.48	22
H5N5F	2012.89	16.86	8
H6N5F t	2174.9	17	4
H6N5F p	2174.9	17	
H7N6F	2540.9	20.05	10
H8N7D	2906.23	23.3	7

Table 1: Results of P4 GPC MALDI MS analysis of CD2 glycans

The products of the digestion of a fraction of the asialo glycan pool with *A.Saitoi* α-1,2 mannosidase is shown in Fig.3b. A series of peaks (7-11) have been reduced to a single peak with a hydrodynamic volume of 8.85gu. This result indicates that CD2 contains an oligomannose series (Man9-Man5) which has been digested to a single population of Man5 glycans. Digestion of a second fraction of the asialo sugars with Jack Bean α-mannosidase reduced the same series of glycans to sugars eluting at 6.8, 6.3 and 5.6gu, elution positions which are consistent with those of Man3, 2 and 1 (Fig. 3c). The data suggest that small oligomannose sugars are relatively resistant to the enzyme. The sugars from the 6.8, 6.3 and 5.6 gu peaks were therefore pooled and digested further with Jack Bean α-1,2,6>3 mannosidase to yield mainly Man1 (Fig. 3d).



3b

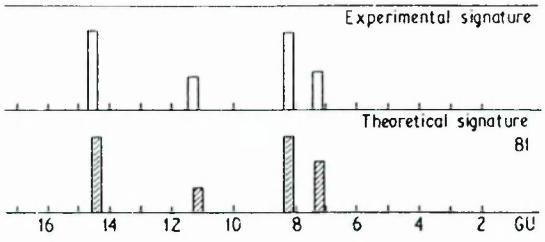


3c

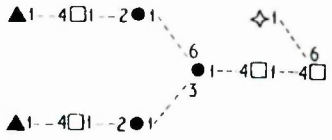
Figure 3b,c: The P4 GPC separation of the CD2 neutral glycans *following digestion with (b) A. Saitoi α1-2 mannosidase (c) Jack Bean α-mannosidase. The hydrodynamic volume of the peak fractions are marked in GU.*

a. Matching signatures – N-Glycans

Database Ref	Degeneracy	Match Quality
comp-7457	1	81

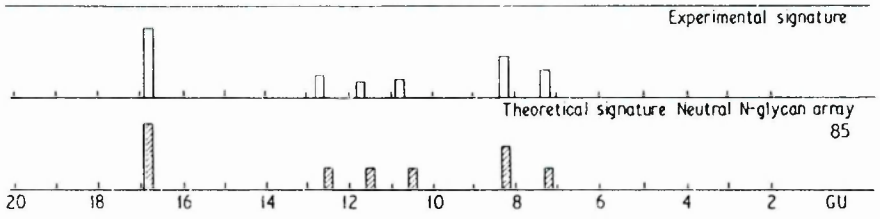


comp-7457 14.4 GU 1788 Da

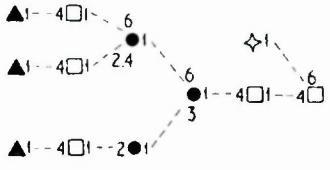


b. Matching signatures – N-Glycans

Database Ref	Degeneracy	Match Quality
comp-7453	2	85

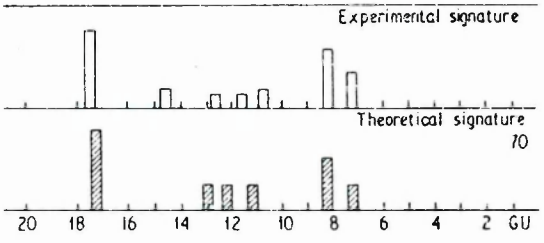


comp-7453 16.8 GU 2153 Da

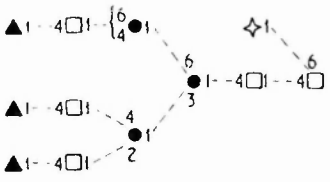


c. Matching signatures – N-Glycans

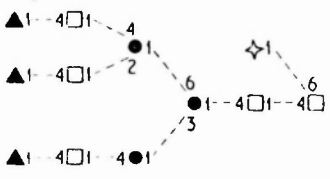
Database Ref.	Degeneracy	Match Quality
comp-6529	3	70
comp-7201	2	68
comp-7453	2	68



comp-6529 17.3 GU 2153 Da

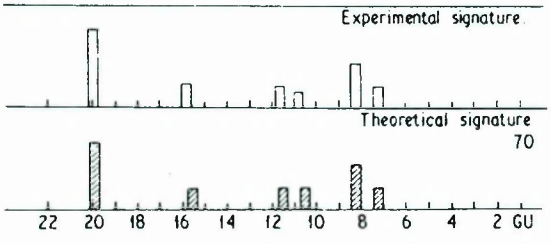


comp-11469 17.3 GU 2153 Da



d. Matching signatures – N-Glycans including polylactosamines

Database Ref.	Degeneracy	Match Quality
poly-131069	6	70
poly-125521	9	67



poly-131069 19.9 GU 2518 Da

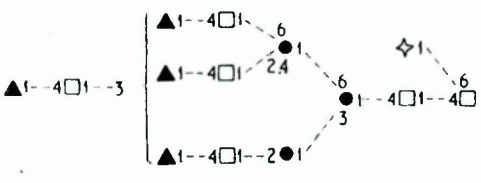


Figure 4 a-d RAAM analysis of asialo glycans from CD2. Peaks 3, 4, 5 and 6 (Fig. 3a) were identified as H7N6F (19.9gu), H6N5F (17.3gu), H6N5F (16.8gu) and H5N4F (14.4gu) respectively.

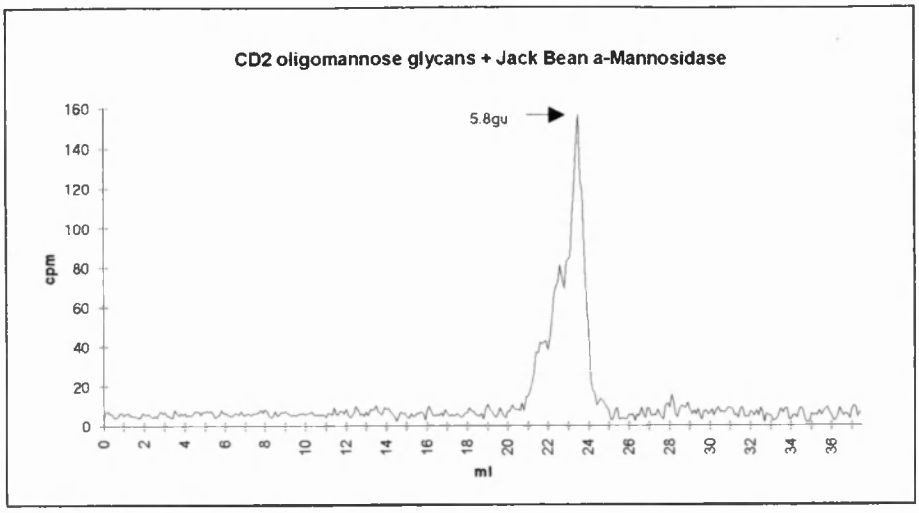


Figure 3d: P4 analysis of the products of the extensive digestion of the glycans at 6.8, 6.3 and 5.6gu (Fig. 3b) with Jack Bean α-mannosidase

The data in figure 3a-d indicate that CD2 contains an oligomannose series consisting of Man9: 3%, Man8: 3%, Man7: 18%, Man6: 12%, Man5: 10% (Tables 1 and 2a, below). Oligomannose glycans account for 46% of the total oligosaccharides associated with CD2.

3.3.3 CD2 does not contain hybrid type structures

A comparison of figures 3a and b indicates that all the glycans from the asialo glycan pool which were digested by *A. Saitoi* mannosidase were reduced to Man 5 suggesting that CD2 does not contain hybrid type sugars carrying a mannose arm susceptible only to Jack Bean α-mannosidase.

3.4 Analysis of individual glycans by Reagent Array Analysis Method (RAAM)

The peak fractions from each of the peaks numbered from 2-6 (along the x-axis) in figure 3a were analysed using the standard RAAM N-glycan array (see ch.8 section 1.8).

(i) Peaks 3, 4, 5 and 6 were identified as H7N6F (19.9gu), H6N5F (17.3gu), H6N5F (16.8gu) and H5N4F (14.4gu) respectively (see figure 4a-d).

(ii) The tri-antennary glycan, H6N5F, was present in at least two isoforms. These were identified by RAAM analysis (4b and 4c) although they were indistinguishable by MS. There is no difference in the molecular weight of these isoforms nor in the number of branches they contain, however differences in arm specificity between the degenerate structures in 4b and 4c gives rise to different molecular volumes, and in the original P4 separation of the asialo CD2 glycans they elute at 16.8 and 17.3 gu respectively, in peaks 4 and 5 (Fig.3a).

(iii) RAAM analysis failed to identify the glycans in peak 2; this may be because the molecular volume of the original glycan exceeded that of the internal standard dextran marker. However the MS data (Fig. 5) suggests the peak (Fig. 3a) contains H8N7F (23.3gu). This result highlights the need to develop an automated sequencing strategy for large polylactosamine structures. A number of glycans all with the structural formula H6N5F were identified within the envelope of peaks 4 and 5 by a combination of RAAM and MS (section 3.6 below). These consist of two sets (Fig. 6), the bi- and tri- antennary isoforms.

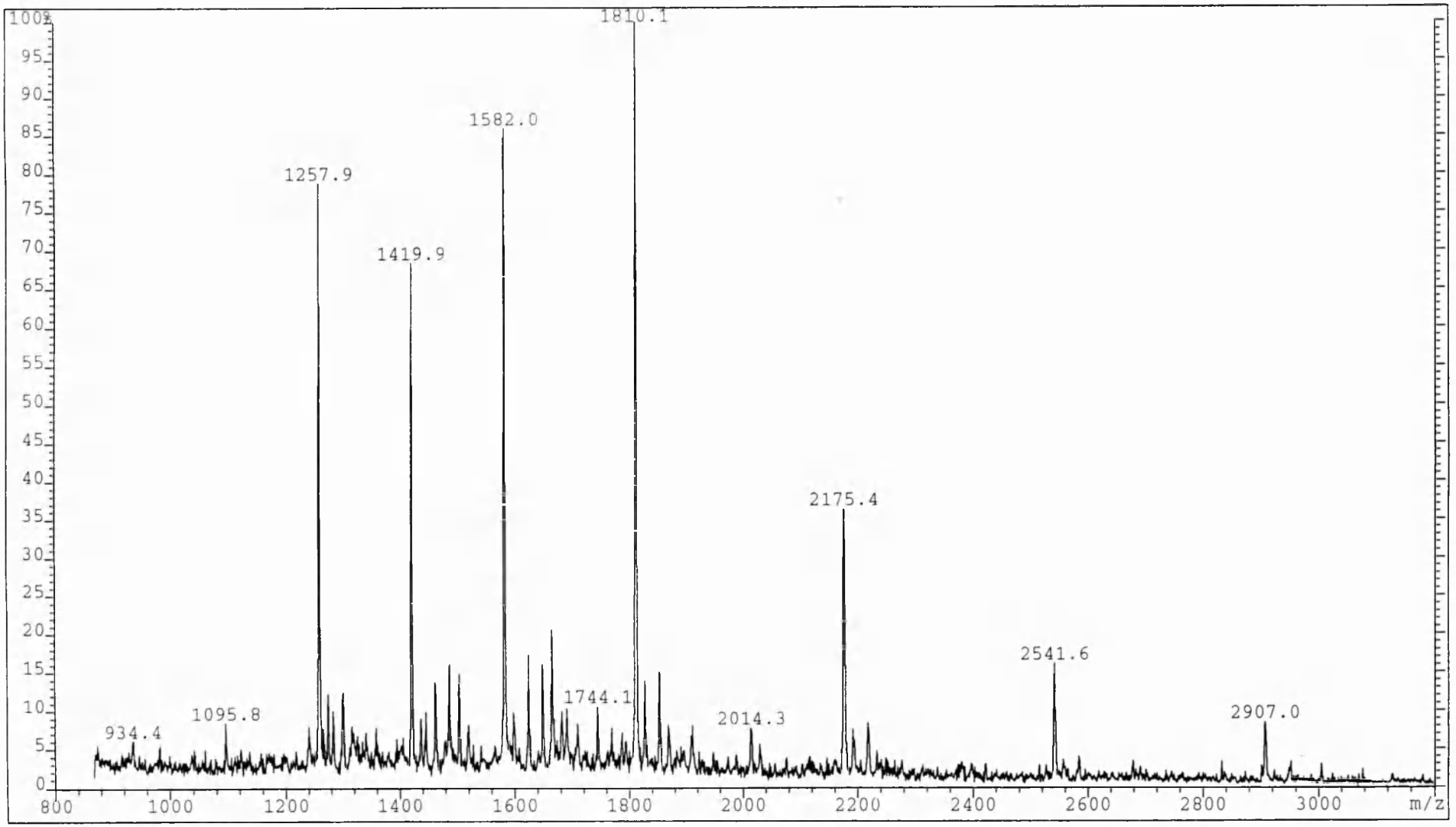


Figure 5 MALDI MS analysis of CD2 asialo glycans. The data are analysed in table 2b

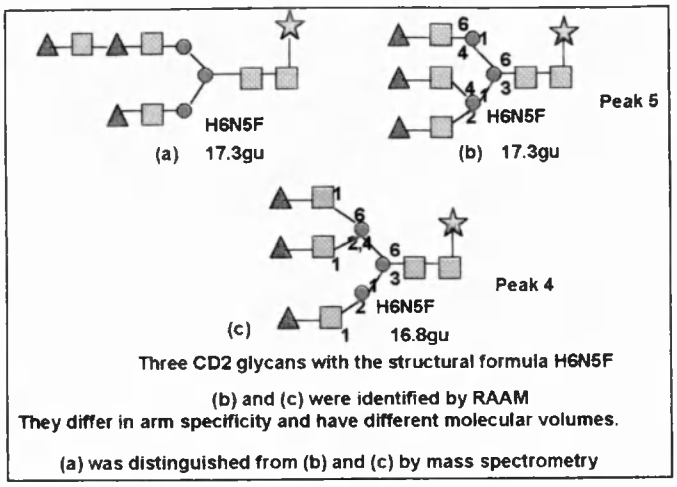


Figure 6

3.5 Analysis of asialo glycans of CD2 by MALDI MS

The desialylated, purified, unlabelled glycans released from CD2 were analysed by MALDI MS (Fig.5, Table 2a). To confirm the finding that CD2 contains an oligomannose series, but not hybrid structures, another fraction of the desialylated CD2 glycan pool (Fig.7a) was extensively digested with Jack Bean α-mannosidase and analysed by MALDI MS (Fig. 7b). The results (Table 2a) are consistent with the P4 data.

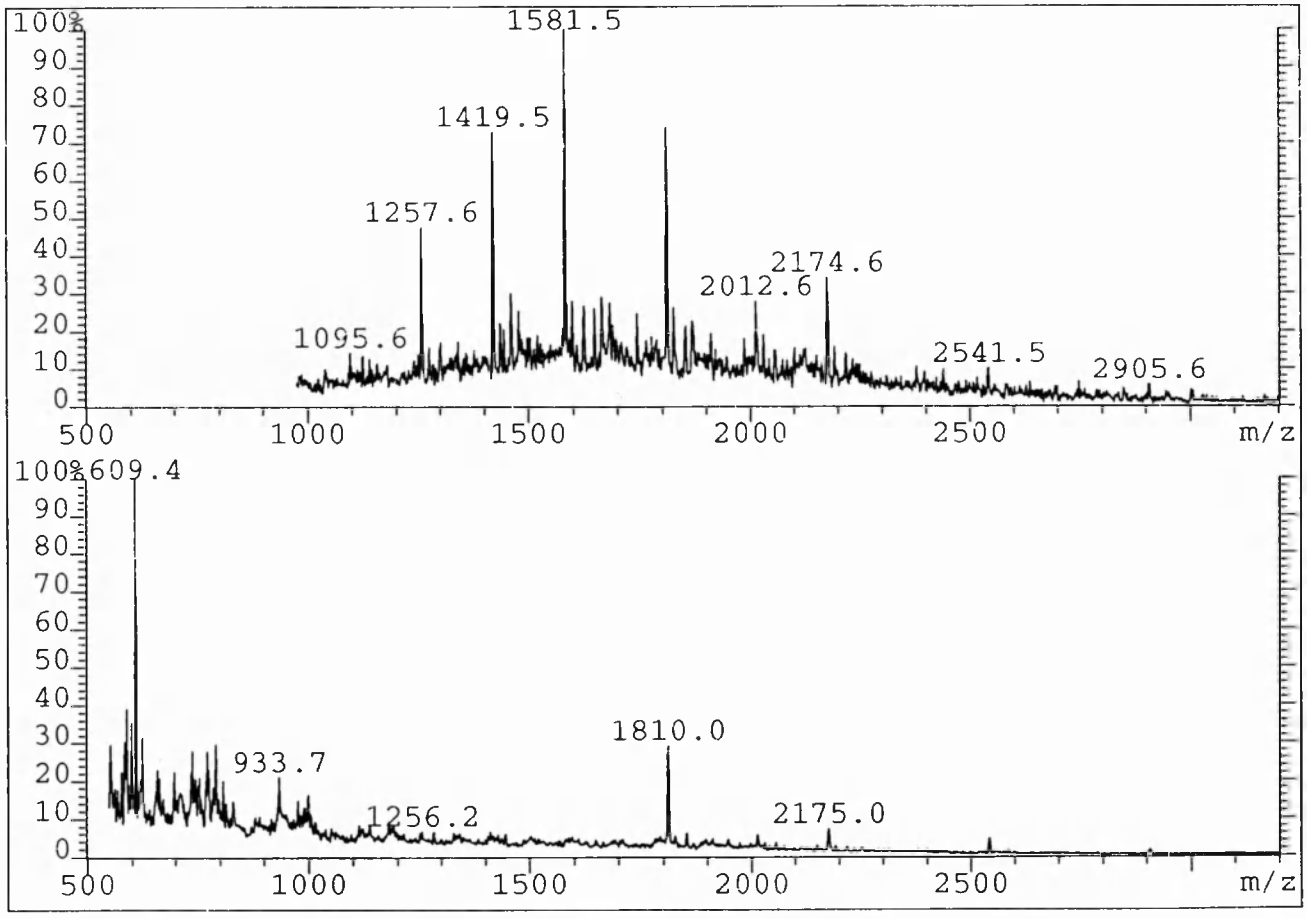


Fig 7a,*b The MALDI MS analysis of the asialo CD2 glycan pool* *purified from protein and salts on a mixed bed resin. The oligomannose series is M4GlcNAc2 - M9GlcNAc2 (7a) with*

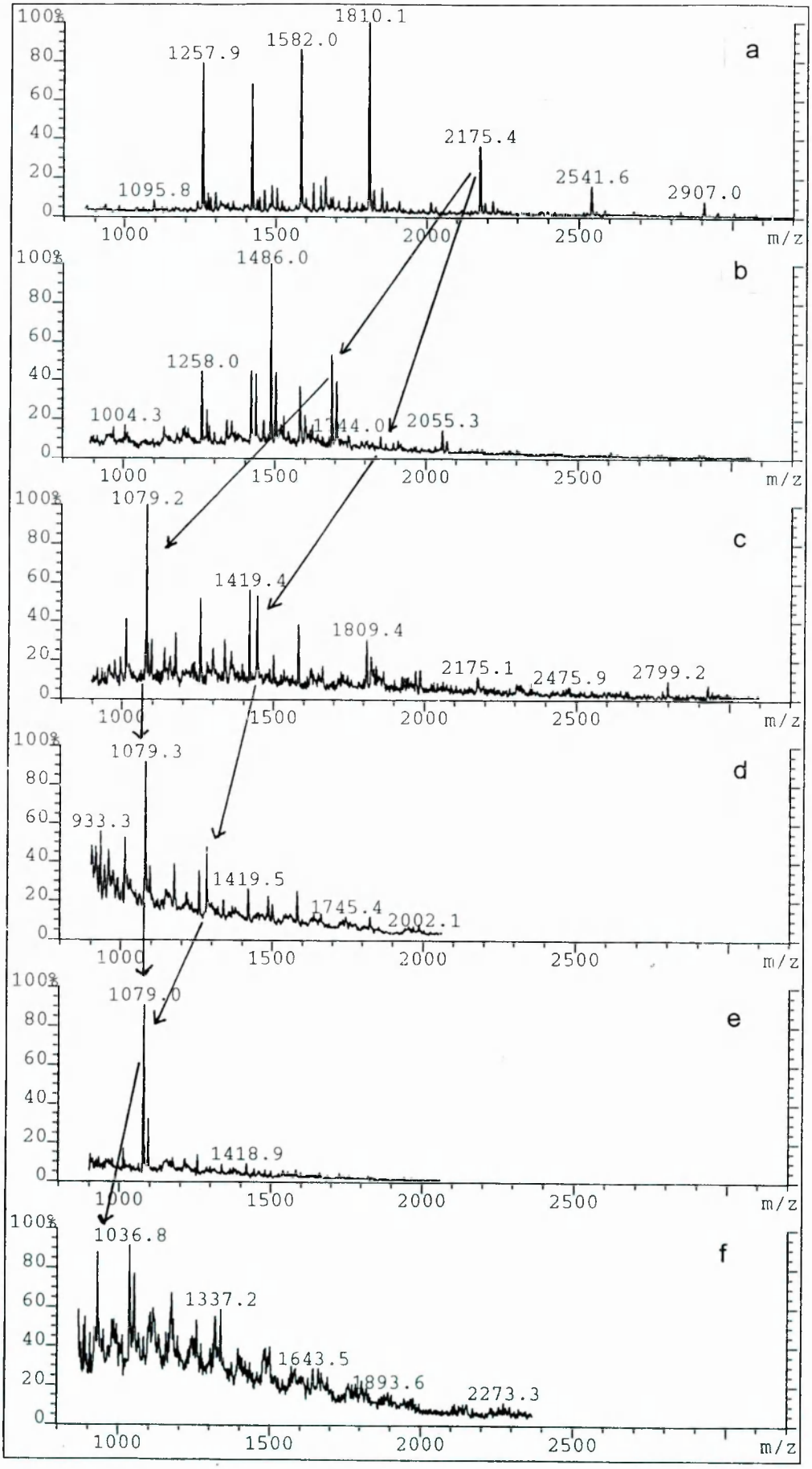


Figure 9a-f: Simultaneous exoglycosidase sequencing of CD2 asialo glycans using MALDI MS. *The arrows show the different sequential digestions of two glycans both with the structural formula H6N5F. The digestions are shown diagramatically in figure 8.*

m/z values of 1095.6, 1257.6, 1419.5, 1581.5, 1743.5 and 1905.4 respectively. After extensive digestion with Jack Bean α-mannosidase the mannose series is reduced to a single population of Man1GlcNAc2 (Fig. 7b) m/z: 609.4.

3.6 Determination of CD2 asialo glycans by simultaneous exoglycosidase digestions and analysis by MALDI MS.

A fraction of the asialo glycan pool (MS fig.9a) was digested sequentially with Bovine Testes β-galactosidase, Jack Bean β-hexoseaminidase, Bovine Testes β-galactosidase, Jack Bean β-hexoseaminidase, and *Choronia lampas* α-fucosidase. Two typical examples of the digestion sequence are shown diagramatically in Figure 8. After each incubation the sample was purified on a mixed bed resin and analysed by MALDI MS (figure 9b-f).

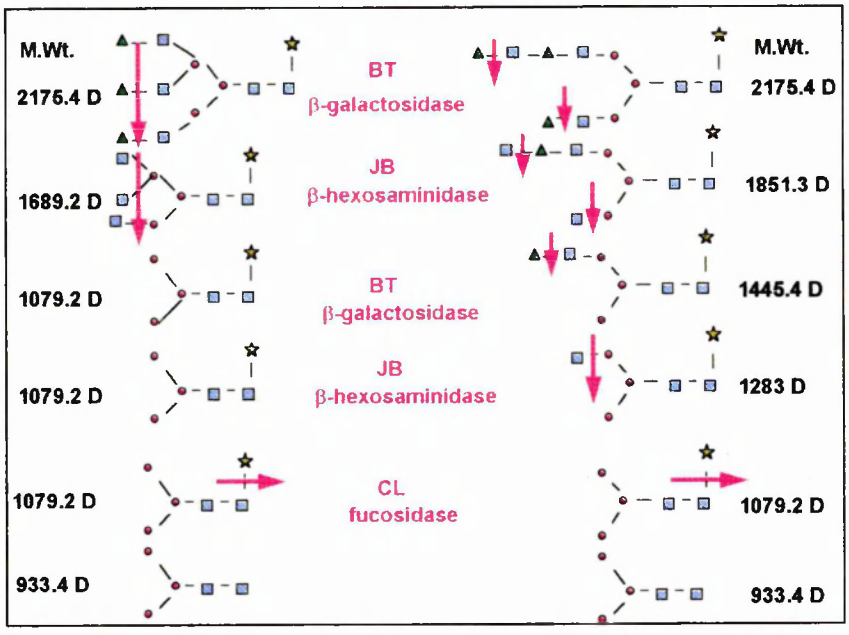


Figure 8: Typical examples of the effects of the exoglycosidase sequencing digestions.

The glycans have been chosen to illustrate the way in which the combination of enzyme digestions and MALDI MS has been used to distinguish between two isoforms of H6N5F, which have the same molecular weight, but which react differently with the sequencing enzymes in this analysis.

The structural formulae of the glycans determined by mass spectrometry were consistent with the P4 analysis and are summarised in Table 2b and figure 10.

Glycan	Mol. Wt.	G.U.	%	JBManGU	JBManMS
M5	1257	8.85	10	5.62	609
M6	1419	9.79	12	5.62	609
M7	1581	10.82	18	5.62	609
M8	1743	12.16	3	5.62	609
M9	1905	13.84	3	5.62	609
H5N4F	1810	14.48	22	14.48	1810
H5N5F	2013	16.86	8	16.86	2013
H6N5F t	2175	17	4	17	2175
H6N5F p	2175	17		17	2175
H7N6F	2540	20.05	10	20.05	2540
H8N7D	2907	23.3	7	23.3	2907

Table 2a

	Ch.mass	a. MS:m/z	b. BT Gal	c. JB Hex	d. BT Gal	e. JB Hex	f. CL Fuc
M5	1257.9	1257.9	1258	1256	1256	1257	1257
M6	1418.9	1419.9	1419.5	1419.4	1419.5	1418.9	1419
M7	1581.6	1582	1582	1581	1581	1581	1581
M8	1743.74	1744.1			1745.4		
M9	1905.88	1905.8					
H5N4F	1809.7	1810.1	1486	1079.2	1079.3	1079	933.4
H5N5F	2012.89	2014.3	1689.2	1079.2	1079.3	1079	933.4
H6N5F t	2174.9	2175.4	1689.2	1079.2	1079.3	1079	933.4
H6N5F p	2174.9	2175.4	1851.3	1445.4	1283	1079	933.4
H7N6F	2540.9	2541.6	2055.3	1445.4	1283	1079	933.4
H8N7F	2906.23	2907	2421	1809.4	1488	1079	933.4

Table 2b

Table 2a: Analysis of data in figure 3a, the P4 analysis of the asialo glycans of CD2.

Table 2b: Analysis of data in figure 5, the MALDI MS analysis of the asialo glycans of CD2 and figure 9, the sequential digestions of the glycans.

The following conclusions can be drawn from the MALDI MS analysis of CD2 asialo glycans:

(i) CD2 contains an oligomannose series (Man5-Man9) and a series of five complex sugars (Fig. 6) which have the general formulae H5N4F, H5N6F, H6N5F, H7N6F and H8N7F. These two series are shown diagramatically in figure 10.

(ii) All the branches of all the complex glycans terminate in galactose. Following digestion with BT galactosidase MS shows that all structures have been reduced in size by the same number of galactose residues as the branches they contain (Fig. 9a,b).

(iii) H6N5F (Fig. 3a: 16.8-17.3gu, Fig. 7a: 2175 mass units) is a mixture of glycans i.e. a triantennary complex structure and a bi-antennary polylactosamine glycan with the same molecular weight. Following digestion with galactosidase, the 2175 mass units peak (fig. 9a) has been reduced to two peaks in Fig. 9b (1689 mass units and 1851 mass units). The digestions of these two species (shown schematically in figure 8) can be followed independently through the remaining profiles (9c-f).

(iv) All the glycans are core fucosylated; this is confirmed by the final digestion with fucosidase (Fig. 9d to 9e) in which all of the structures accumulated at 1079.3 mass units were reduced by approximately 145.9 to 933.4 mass units.

(v) The oligomannose series is unaffected by all of these enzymes. The relative amounts of these structures appears to decrease and there have been some losses of material; in addition, since the complex sugars accumulate into a smaller number of peaks following each digestion, the mannose glycans appear less by comparison.

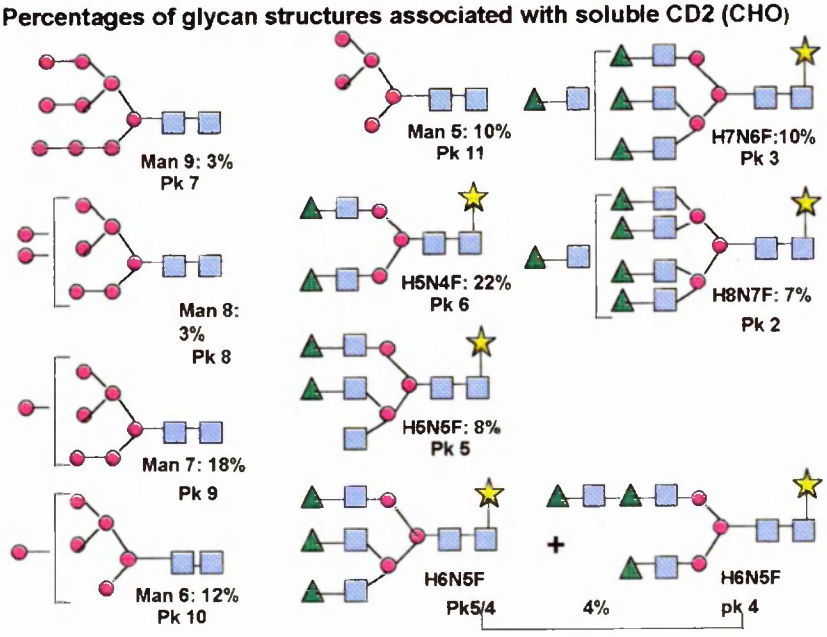


Figure 10: Proposed structures of N-linked glycans associated with soluble CD2

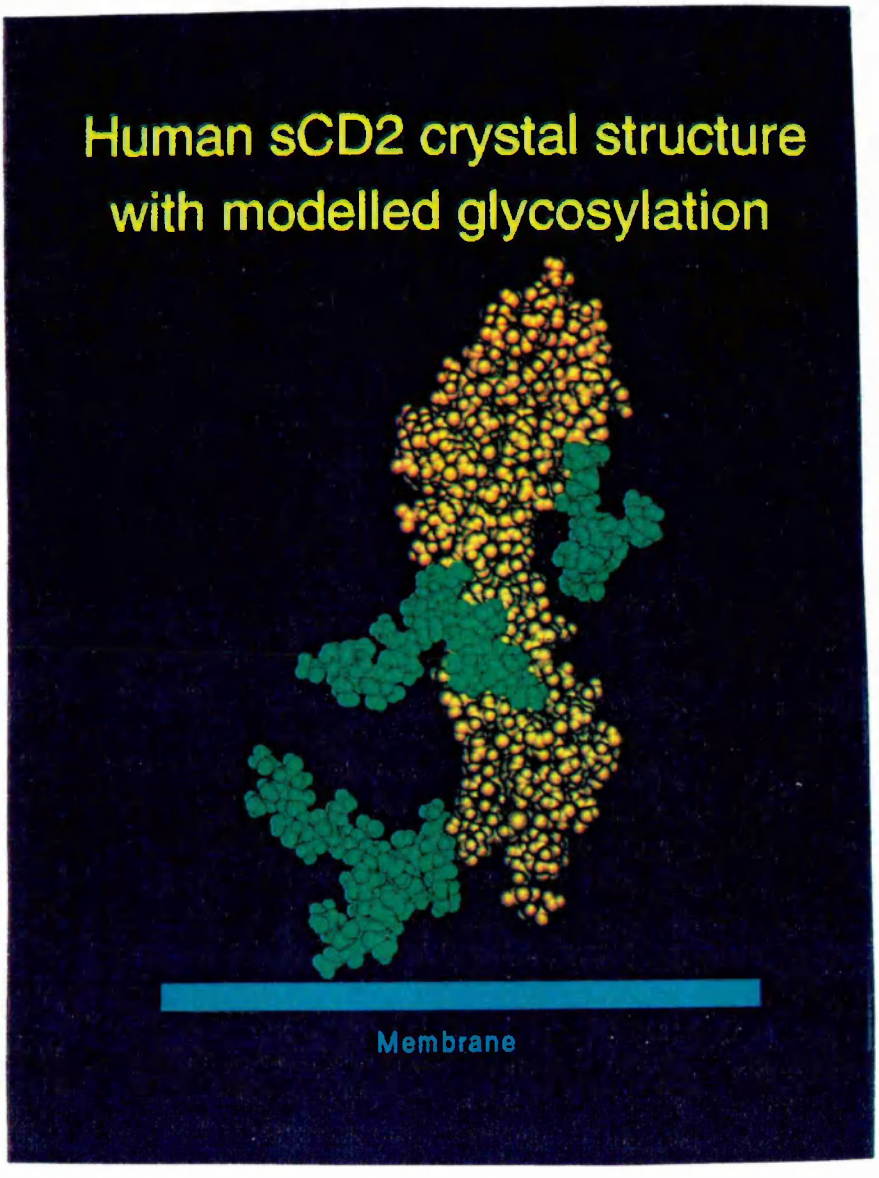


Figure 11: The leucocyte antigen CD2 dimer *modelled with the Man6 glycan at site 65, the disialylated biantennary core fucosylated complex sugar at site 126 (H5N4F), and a triantennary structure with a lactosamine extension at site 117 (H7N6F). The binding site for CD58 (LFA3) on monomeric CD2 is close to the N-terminus on the opposite face of the molecule to the glycans at site 65.*

3.7 A molecular model of glycosylated CD2 has been proposed.

A molecular model was constructed based on the X-ray crystal structure of human CD2 (Bodian et al 1994) and the glycan analysis from this study. The site specificity of the glycans has yet to be established, however glycan analysis of domain 1 (Recny et al 1990) has allowed us to propose a model of one of the possible glycoforms of CD2 (Fig. 11). The glycans at each site were selected on the following basis:

(i) *Site Asn65GlyThr*

46% of the glycans attached to CD2 were found to be of the oligomannose type. MS analysis (Recny et al 1992) of the glycans attached to domain 1 at Asn 65 found only the oligomannose type. Taken together these data suggest that site 65 should contain only oligomannose glycans of which the greatest proportion in the analysis were of the Man7 type (39%). Site 65 has been modelled with a Man6 sugar (12% of total glycans) attached to the Asn. The N-glycans at Asn 65 are attached to a tight loop connecting β-strands D and E of a classical Ig β-sandwich domain (Jones et al 1992). It has been proposed that, depending on the length and flexibility of the D/E loop and the orientation of the asparaginyl side chain within this loop, the oligomannose glycans at Asn65 may be positioned alongside β-strands B, D and E in the face of the β–sheet opposite to the proposed binding site for CD58. From the molecular model (Fig.11) it appears that the sugars can fit into a cleft in the β-strands of the protein. However NMR spectroscopy studies by Wyss et al (1995) found no interactions between the protein and the sugar, suggesting that the glycans are not situated in the cleft.

(ii) Site Asn126ThrThr

The two domains of CD2 are linked by 6 amino acids which are relatively close to this glycosylation site. 50% of all the structures at sites 117 and 126 consist of the biantennary core fucosylated complex sugar, H5N4F (Fig.10). This predominant structure was modelled at site 126.

(iii) Site Asn117GlyThr

Asn 117 is close to the membrane surface. By analogy with the interaction of CD59 with C5b-8 (see ch.3) the N-glycans at this site may orient the binding site towards the receptor ligands in CD58 and CD48. Over 50% of the glycoforms of CD59 (one glycosylation site) are large polylactosamine structures, therefore the tri-antennary glycan which contains a lactosamine extension, H7N6F (Figure 10; 23% of structures at sites 117 and 126), has been modelled at site 117.

3.8 Analysis of the asialo N-linked glycans associated with rat CD48, a low affinity ligand of rat CD2, by MALDI MS

The glycans from the six N-linked glycosylation sites on rat CD48, expressed in CHO cells, were released by hydrazinolysis, desialylated with *Arthrobacter ureafaciens neuraminidase* (fig.2a-c), purified on mixed bed resins and analysed by MALDI MS. At least 26 different glycans were detected, without accounting for isomers (Fig 12, Table 3). Figure 12 has been

M5 1257	**H3N3F** 1282.42	**H3N4F** 1485.5	**H3N5F** 1688.58		
	H4N3F 1444.4	**H4N4F** 1650.1	**H4N5F** 1852.8		
		H5N4F 1810.7	**H5N5F** 2012.5	**H5N6F** 2219.4	**H5N7F** 2422.5
			H6N5F 2177.5	**H6N6F** 2378.6	**H6N7F** 2585.5
				H7N6F 2543.7	**H8N7F** 2908.7

Table 3a: Composition of fucosylated glycans identified by MALDI MS in the pool of oligosaccharides released from CD48.

H3N3 1136.36	**H3N4** 1339.44			
H4N3 1298.41	**H4N4** 1501.5	**H4N5** 1704.6		
	H5N4 1663.5	**H5N5** 1866.6	**H5N6** 2069.7	
			H6N6 2231.73	
			H7N6 2393.8	

Table 3b: Composition of non-fucosylated glycans identified by MALDI MS in the pool of oligosaccharides released from CD48.

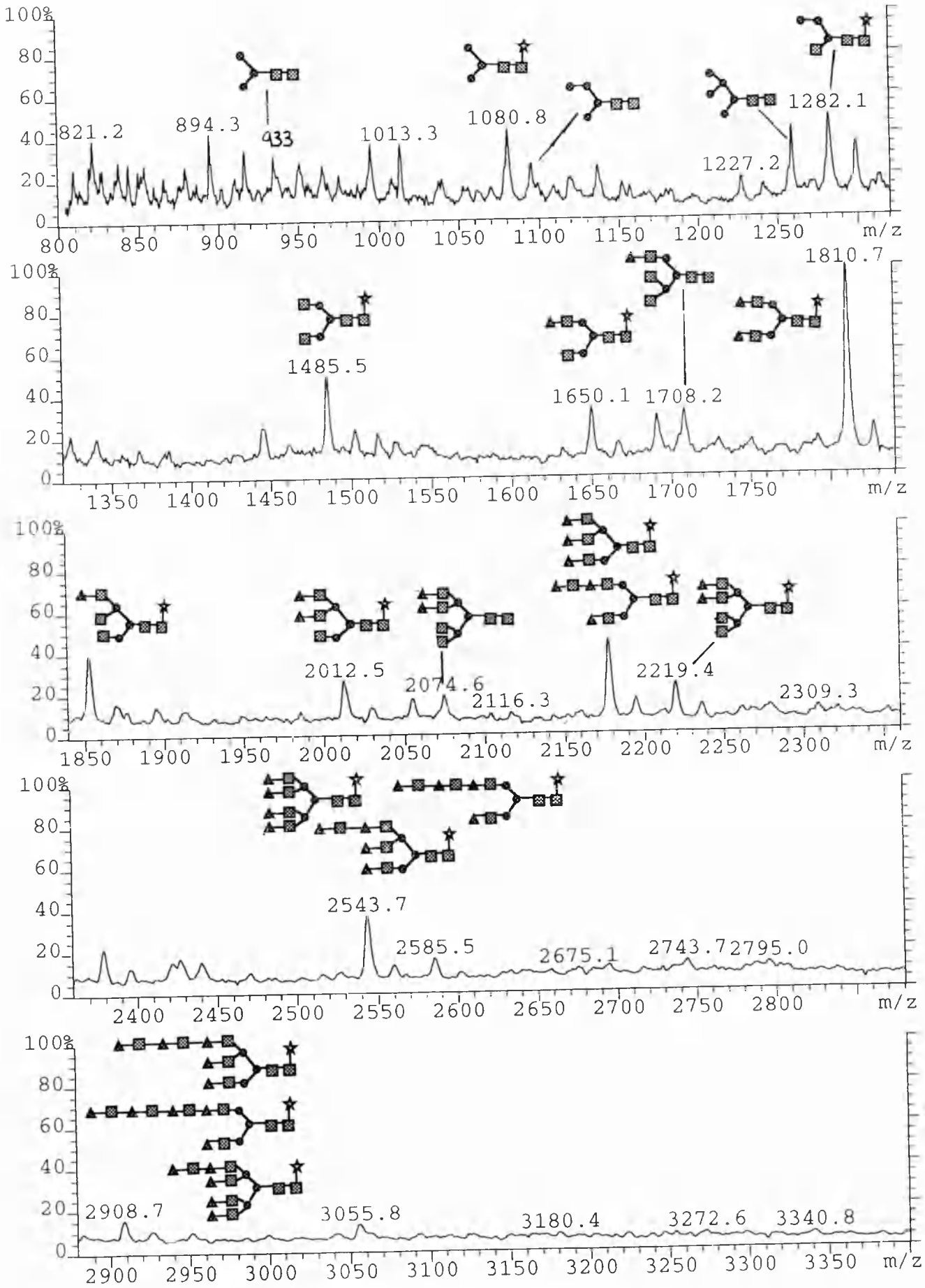


Figure 12: The MALDI MS analysis of CD48
The data are analysed in Tables 2a and 2b.

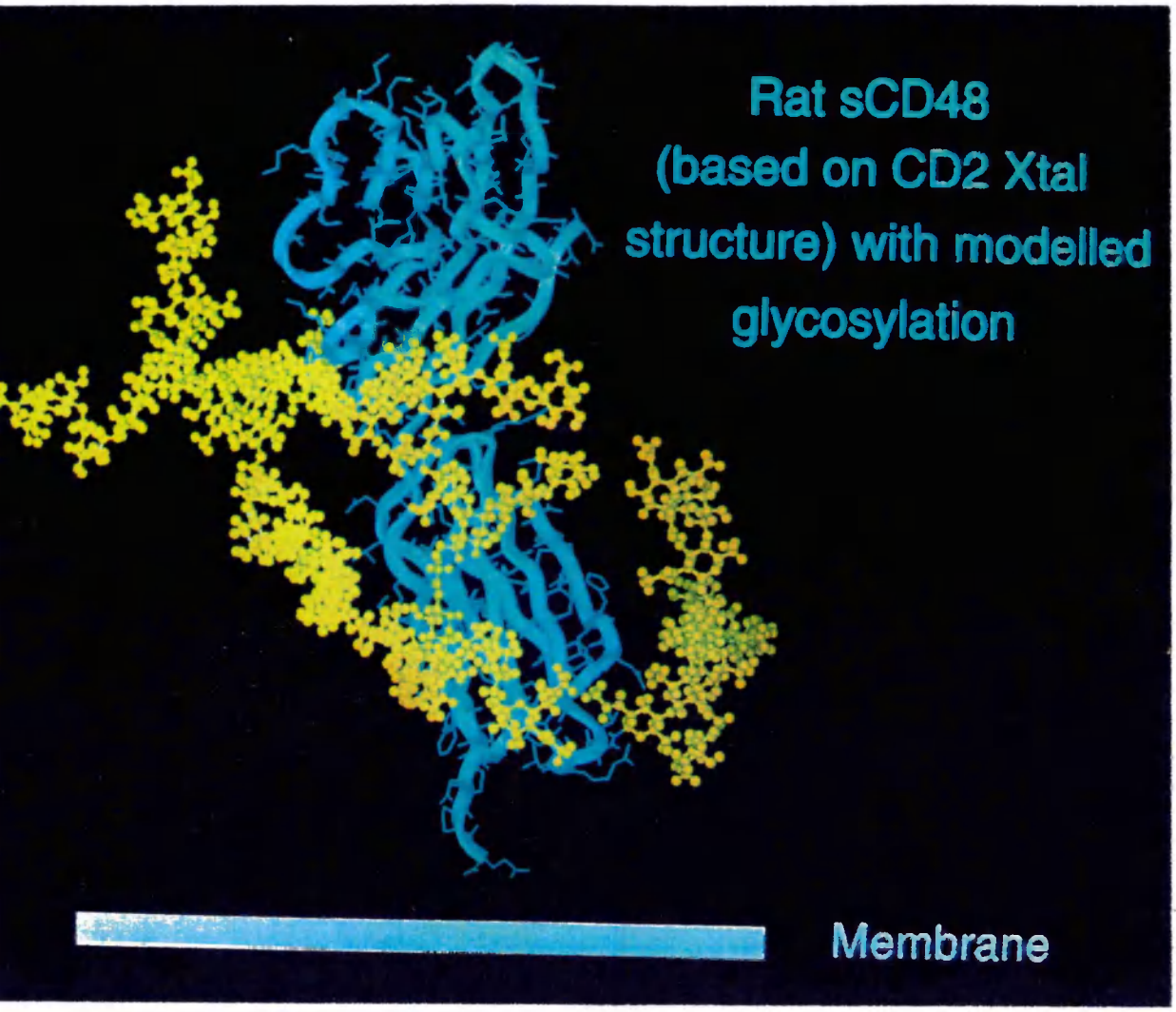


Figure 13: Rat soluble CD48 with modelled glycosylation.

The crystal structure of CD48 is not available, therefore the protein structure (blue ribbon) has been based on the CD2 crystal structure (Bodian et al 1994). Sequence alignment has been used to obtain equivalent residues in CD2 for the oligosaccharide substitutions in CD48. CD48 is glycosylated at sites at Asn 16, 75, 164 and 181. The equivalent positions in CD2 are 13, 67, 150 and 167. The site distribution of the oligosaccharides is not known. The oligosaccharides which are modelled at each of the sites (yellow ball and stick) are selected from the most prevalent structures identified in this analysis. Sites 150 and 167 contain the biantennary structure H5N4F, site 13 contains SAH6N5F (a monosialylated biantennary sugar with one polylactosamine repeat) and site 67 contains the tetra antennary sugar, H7N6F.

annotated with some possible alternative structures which are consistent with the MS composition analysis. The major structure (1808 mass units) had the same mass/ charge ratio as that predicted for the fucosylated bi-antennary glycan H5N4F. The largest structure detected was 3038 mass units, consistent with the general formula H8N7F. The smallest was a glycan of 933 mass units (Man 3). There were two series of oligosaccharides. The major series was fucosylated while the minor one contained most of the same structures without fucosylation. The origin of the oligomannose glycans smaller than Man 5 is unclear since these are not normal biosynthetic products.

The approximate proportions of the different populations of glycan classes was determined by measuring peak height. A major advantage of the magnetic sector instrument used in this MALDI MS analysis is that the individual peak heights are directly proportional to the ion concentration of each component. It has also been demonstrated that oligosaccharides are non-selectively ionised (Dr. D. Harvey - unpublished data). However, to obtain quantitative data it is necessary to take many scans (typically 20-30) over a wide area of the sample spot, since accurate quantitation depends on an even distribution of the individual sugars across the target area. MS also gives only the m/z ratios of the oligosaccharides and no linkage information is available from these data.

The relative proportions of the different classes of glycans in CD48 based on peak height measurements were approximately as follows: oligomannose 10%, hybrid 10%, complex biantennary 40%, and polylactosamine complex glycans 20%. The remaining structures have not been identified and may be fragment ions.

4. Conclusions

The MALDI MS analysis of the four glycans attached to rat sCD48 suggests that processing by the CHO cell glycosylation machinery has terminated at almost every possible stage. At least 47 different glycans are N-linked to CD48. Interestingly the molecular model indicates that all four sites are on the same face of the molecule and this may suggest that there may be limited access to some of the processing enzymes, in particular the sialyl transferase, and that consequently glycosylation processing at one site affects another. This is in contrast to CD4 where glycosylation at one site did not affect the processing at others (Ashford et al 1993). 90% of the glycans attached to CD2 were neutral, and 10% were monosialylated. All the charge was due to 1,3-linked sialic acid. An oligomannose series (Man 9-Man 5) and five different core fucosylated, fully galactosylated, complex glycans ranging in size from H5N4F to H8N7F were identified. The heterogeneity of the sugars attached to this protein appears to be very extensive. The identification of three distinct glycans all with a general formula H6N5F (two isoforms of a tri-antennary structure and an extended polylactosamine bi-antennary glycan) beneath the envelope of peaks 4 and 5 in the P4 separation (figure 3a) indicates the need for second dimensional methods for separating structures co-eluting on GPC. This study also demonstrates the use of a combination of P4 GPC and MS for the analysis of glycan mixtures. 46% of the total glycan

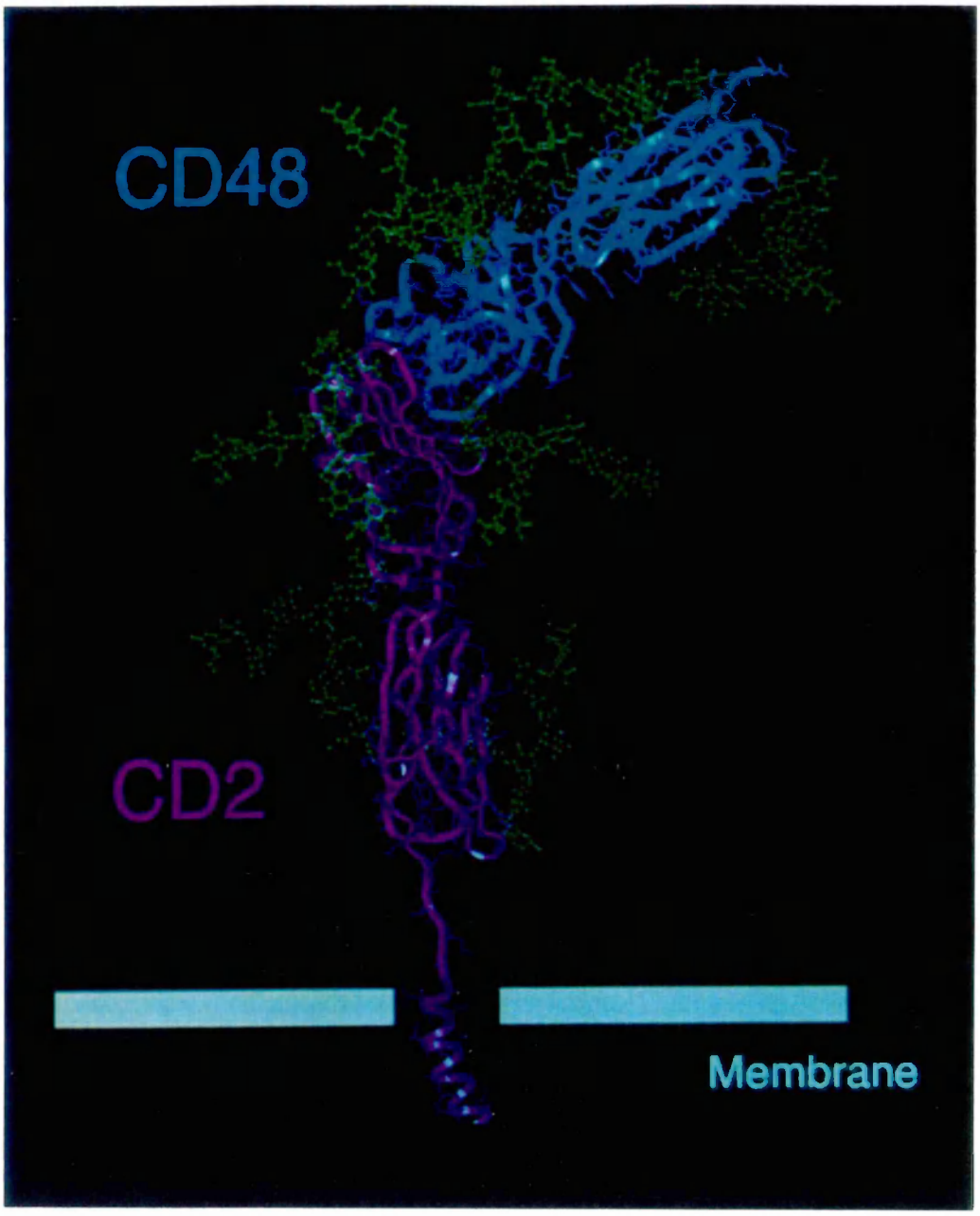


Figure 14: The interaction of rat CD2 and rat CD48 with modelled glycosylation. *The protein structure of CD48 is based on the CD2 crystal structure (Bodian et al 1994; see figures 11 and 13). The glycosylation of CD48 is based on data from this thesis and the glycosylation of rat CD2 on data taken from Davis et al 1993). The topology of the CD2/CD48 interaction is based on the study by van der Merwe et al (1995).*

pool consists of an oligomannose series, Man 5-9. Assuming that all the glycosylation sites on CD4 are fully glycosylated and that site 65 contains only oligomannose sugars (Recny et al 1992) this implies that 13% of the glycans at sites 117 and 128 are also of the oligomannose type.

The sugars shield a large area of the protein surface and may give protection from proteases as has been demonstrated for ribonuclease (ch.4). A possible structural role for each of the oligosaccharides was explored using a molecular model of glycosylated CD2 (Fig.11).

(i) *Site Asn65GlyThr (domain 1)*

From the molecular model (Fig.11) it appears that the sugars can fit into a cleft in the β-strands of the protein and stabilise a particular conformation of the molecule. However NMR spectroscopy studies by Wyss et al (1995) found no interactions between the protein and the sugar, suggesting that the glycans do not lie within this cleft. However the local protein conformation at the cleft may shield the developing glycan chain from the biosynthetic enzymes which process oligomannose structures to hybrid and complex glycans, ensuring that the glycan processing terminates at the oligomannose stage. Interestingly, while the sugars are not in a position to serve as a ligand for the receptor, both de-N-glycosylated (PNGase F) sCD2 as well as sCD2 in which the site at Asn 65 has been deleted fail to bind CD58. (Recny et al 1992). Activity is retained when the oligomannose N-glycans are trimmed back to $Man_3GlcNAc_2$ by digestion with α-mannosidase indicating that a minimum N-glycan core tri-saccharide is required to give conformational stability within the CD2 adhesion domain and maintain a functional CD58 receptor binding site. (Knoppers et al 1992). Functional activity was also retained when the protein was expressed in the presence of a glycosidase inhibitor, BuDNJ, which leaves unprocessed $Man_9GlcNAc_2Glc_3$ residues at each glycosylation site (Davis et al 1995).

(ii)Site Asn126GlyThr (Domain 2)

The two domains of CD2 are joined by a link region of 6 amino acids (99-104). The molecular volume of the glycans at site 126 is large compared with the protein and this raises the possibility that the sugars at this site may interfere with the orientation or rotation of the two domains. In general membrane anchors restrict the rotation and translation of cell adhesion molecules; in CD2 the linker region is a key factor which allows additional freedom to domain 1, which carries the binding site. interestingly, the major difference between rat (which does not have a sugar at this site) and human CD2 is a 13-20° difference in the orientation of the two domains observed in the crystal structure (Bodian et al 1994).

(iii) Site Asn117ThrThr (Domain 2)

Asn 117 is close to the membrane surface and is the only site conserved between rat and human CD2. The glycosylation sites in rat CD2 are at Asn 67, 77, 84 and 112. By analogy with the interaction of CD59 with C5b-8 (see ch.3) the N-glycans at this site may interact with the membrane protecting the protein from interactions with the lipid bi-layer. The large

size may be a means of influencing the spacing of cell surface molecules. In addition it may orient the binding site towards the receptor ligands in CD58 and CD48 by reducing the space available to domain 2 of the protein.

1. The structural diversity of the leucocyte antigens in this study

The recombinant glycoproteins studied in this chapter have different structural features and levels of complexity (Table 1a). CD59 is a member of the Ly-6 family, CD5 contains a scavenger receptor domain while CD48, CD4 and CD2 are members of the immunoglobulin superfamily. sCD59 has only one domain while the CD5/CD4d3,4 chimera has five. The number of glycosylation sites associated with each protein ranges from one on sCD59 to four on CD48. Two of the glycans in CD2 are attached to the IgSF domains, in CD48 two are attached to the IgSF domains and two are located close to the junctions of the domains and the flexible loops, and in CD4 one is attached to the IgSF domain (see introduction to ch. 2 fig.1).

Antigen	*Domains*	*N-glycosylation sites*	*Ligands*	*Structural studies*
Hu sCD2 (T11, LFA2)	2D	Asn65,117,126	CD58, CD48	Xtal: (Bodian, 1994)
Rat sCD2	2D	Asn67,77,84,112	CD48	Xtal: (Jones, 1992) NMR:(Driscoll, 1991)
Rat sCD48	2D	Asn13,67,150,167	CD2	mutagenesis: rat sCD48/rat sCD2 van der Merwe, 1995
Hu ECD59	1D	Asn18	CD5b-8/C9	NMR (Kieffer, 1994)
Hu CD5d1	1D	Asn92	CD5:CD72	No details available
HuCD5/ ratCD3,4	5D	Asn92 Asn159,270		d3,4 X-ray (Brady, 1993)

Table 1a: Some structural details of the molecules involved in this study

Molecular modelling has been used to explore possible roles for the oligosaccharides. In particular the relative sizes and location of the glycans has been examined (ch.2A section 3.7 and ch.2C section 3.7). The protein structures were taken from X-ray crystallographic or NMR structural data which is available for sCD59 and sCD2. The length of the sugars attached to sCD59 is of the same order of magnitude as the diameter of the globular protein (4.2nm) and are located on the opposite side of the molecule to the binding site for C5b-8/C9. The sugars associated with CD2 are also of the same order of magnitude as the protein domains and none of the sugars appears to be in a position to shield the binding site for CD58. CD2 shows site specific glycosylation and only oligomannose glycans are located at Asn65 (Recny et al 1992) which is located at a cleft in the protein which may restrict access to GlcNAc transferase I. The sugar attached at the flexible link region (Asn117) in CD2 may be in a position to affect the orientation of the two domains. In addition the glycan at Asn126 is close to the membrane, and

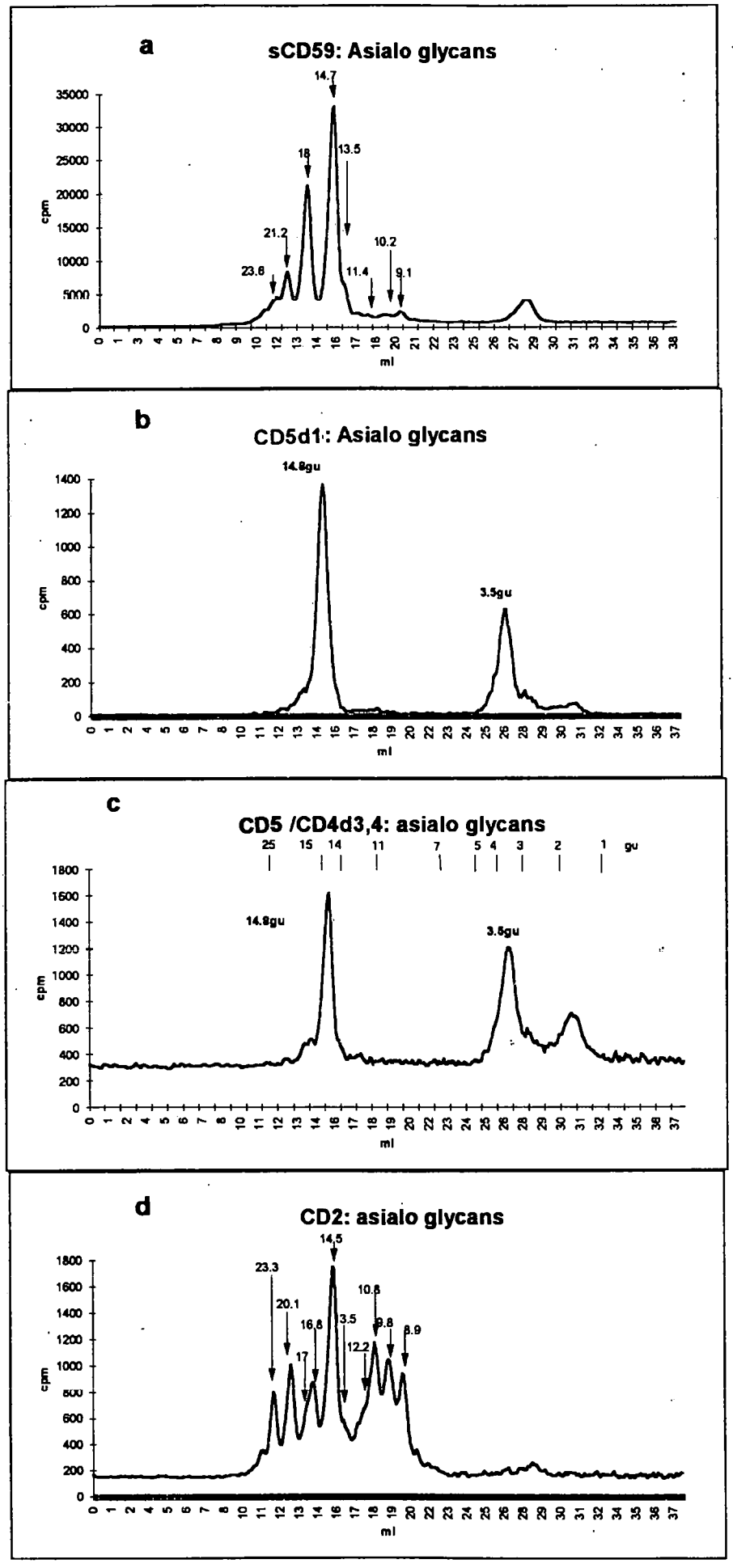


Figure 1: The P4 GPC elution profiles of the asialo glycans released from the soluble forms of the leucocyte antigens studied in this chapter.

(a) CD59 (b) CD5 domain 1 (c) CD5/CD4 domains 3,4 chimera (d) CD2

may, by restricting the space available to the protein and limiting its interactions with the membrane, orient the CD2 binding site towards its ligand CD58. Interestingly, on CD48 the sugars are all located on one face of the molecule, and, as with CD2 and CD59, are not close to the binding site. No structural information is available for the scavenger domains on CD5, which therefore could not be modelled.

2. Processing of complex glycans by CHO cells

The P4 profiles of the asialo-glycans associated with the leucocyte antigens studied in this thesis are shown in figure 1 and the data are summarised in table 1b. The main findings were as follows: (i) the percentage of charged structures (all α2,3-linked sialic acid) ranged from 10% in CD2 to 100% in CD5/CD4d3,4 (ii) the percentage of polylactosamine/multiantennary structures ranged from 45% on sCD59 to 0% on CD5d1 (iii) CD5, very remarkably, contained only one glycoform, the biantennary, core fucosylated complex glycan H4N5F, while at the other extreme CD48 was associated with at least 47 different sugars (iv) only CD2 contained significant levels of oligomannose sugars (v) only CD5d1 and the chimera contained O-linked glycans (v) shorter oligosaccharides were more often sialylated than those containing more than one unit of lactosamine.

Glyco-protein	Doms	No. N-glycans	Charge status				Types of N- glycan			
			%N	A1	A2	A3	%Ol	H	P/M	CB
sCD59	1	1	35	35	12 (b) 18 (t)	0	6	0	45	49
CD5d1	1	1	17	38	45	0	0	0	0	100
CD5/ CD4d3,4	3	2	0	35	65	0	0	0	5	95
CD2	2	3	90	10	0	0	46	0	32	22
CD48	2	4	Not determined				10*	10*	20*	40*

Table 1b Summary of glycosylation analyses in chapter 2

*N-neutral glycans; A1 - glycans containing one sialic acid residue; A2 - glycans containing two sialic acid residues; A3 - glycans containing three sialic acid residues. O-oligomannose; H-hybrid; CB-complex biantennary; PM- polylactosamine and/or multiantennary type glycans; b biantennry, t-triantennary structures. * The data for soluble CD59, CD5 domain 1, the chimera*

and CD2, was derived from integration of the peaks obtained by fractionation of the released oligosaccharides on P4 GPC or WAX chromatography. In contrast the values for CD48 were obtained from the MALDI MS analysis by measuring the peak heights of the various components of the glycan pool. The numbers for CD48 should therefore be regarded only as approximate estimations (see ch.8 section 2.1).

The majority of the glycans associated with these proteins were of the complex biantennary core fucosylated type; the only hybrid structures were identified on rat sCD48, which was found to contain at least 47 possible glycan structures. CD5d1, CD5/CD4d3,4 and sCD59 contain very low levels of oligomannose sugars (6%). Although CD2 contains 46% of oligomannose structures, 33% are at one glycosylation site (Recny et al 1992). These data indicate that, with the exception of CD2 site 65, CHO cells are able to process complex oligosaccharides indicating that the local protein structure allows GlcNAc transferases I and II to transfer N-acetylglucosamine efficiently to the $Man_5GlcNAc_2$ and $Man_3GlcNAc_3$ glycoforms respectively. These are the key enzyme steps in the processing of complex sugars. In CD2, the presence of oligomannose sugars at Asn 65 suggests that GlcNAc transferase I and UDP-GlcNAc may not form an effective ternary complex at this site. Asn 65 is located at a cleft in the protein and this may interfere with complex formation, suggesting that the protein is able to control its own glycosylation in a site specific manner by means of the local protein structure. Interestingly the range of glycosylation in human sCD2 is not reflected in rat sCD2 (4 glycosylation sites) also expressed in CHO cells: the asialo glycans associated with rat CD2 contained 80% of the core fucosylated biantennary complex glycan H5N4F (Davis et al 1993).

2.1 The protein may restrict the number of glycoforms

Domain 1 of CD5 contains one N-linked glycosylation site. The glycans associated with the site were found to be highly restricted, both in size and type. Only a single glycan, the biantennary core fucosylated structure, H5N4F, (83% sialylated), was identified. The absence of microheterogeneity is extremely unusual, therefore to investigate whether the presence of additional domains would increase the range of sugars the glycosylation of a chimera consisting of human CD5/rat CD4d3&4 (5 domains, 3 sites) was analysed. Interestingly, 95% of the glycans were the same complex sugar structure as CD5d1 and 100% were sialylated. The addition of four domains to CD5d1 in CD5/CD4d3/4 therefore resulted in very little extra (5%) processing of the bi-antennary complex glycan which occupied the 100% of the glycosylation sites in CD5d1, suggesting that in this case the quaternary structure does not significantly affect the asialo glycosylation pattern. The glycans associated with rat sCD4 (expressed in CHO cells) is more typically heterogeneous, and include 42% of hybrid and 22% of oligomannose structures and only 14% of H5N4F (Ashford et al 1993). This suggests that the structure of the CD5 scavenger domain may influence the glycosylation processing of both CD5d1 and the chimera. In the construct the scavenger domains of CD5 replace the immunoglobulin domains 1 and 2 which are normally attached to CD4; the glycans are not elongated and this may result

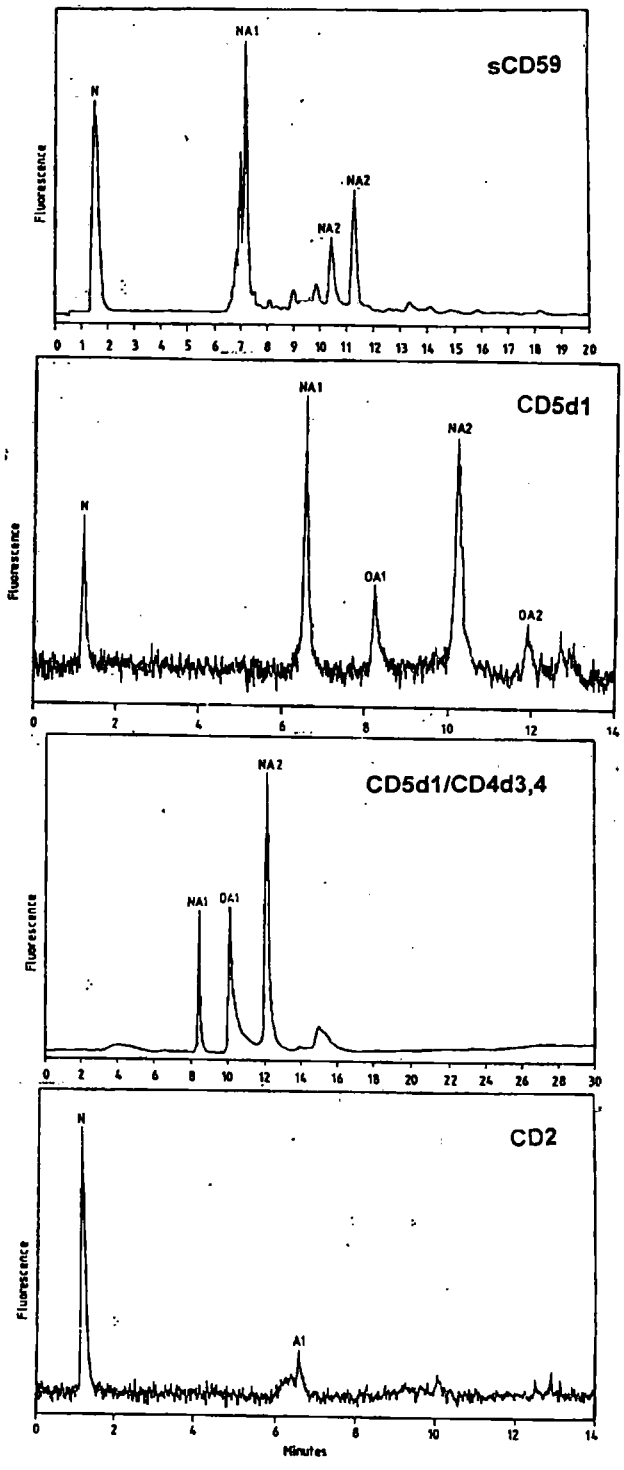


Figure 2: Charge profiles of CHO cell derived CD antigens

from a decrease in the Golgi transit time of the chimera compared to intact CD4. An increased transit time has been shown to increase the percentage of elongated complex glycans (Wang et al 1995). Furthermore, the addition of terminal sialic acid precludes elongation; the high degree of sialylation and the lack of elongation suggests that, compared with outer arm GlcNAc tranferase, sialyl transferase has a high affinity for the complex H5N4F glycans attached to both CD5d1 and the chimera.

2.2 Elongation and branching of oligosaccharides

sCD59 contains a higher proportion of polylactosamine and multiantennary structures (36%) than CD2, CD5d1 and CD5/CD4d3,4. This indicates that the protein structure allows access to the GlcNAc transferases IV, V and VI. Additions of N-acetyl glucosamine by these enzymes is required for a high degree of branching. The presence of a high proportion of polylactosamine and multiantennary structures also suggests that GlcNAc transferase competes effectively with sialyl transferase for outer arm galactose and also that the transit time of the protein through the Golgi is sufficient to allow extension of the bi-antennary complex glycans. Sialylation prevents the addition of lactosamine extensions in approximately 10% of dislalylated biantennary structures. Interestingly, only 12% of sCD 59 glycans are elongated, while 24% are of the triantennary type, indicating that the protein structure does not preclude either possibility. Interestingly, in human erythrocyte CD59 approximately 40% of the glycans are of the polylactosamine type (ch.3) of which 30% are greater than 23gu. This suggests that an important factor in determining the percentage of polylactosamine structures may be the attachment of a GPI anchor which increases the transit time of the glycoprotein through the Golgi. In contrast to the native CD59, human sCD59 expressed in CHO cells does not contain significant amounts of polylactosamine or multiantennary structures larger than 23 gu. Interestingly both glycosylated and unglcosylated sCD59 is expressed and secreted (ch. 2A, Fig. 1b). In contrast, human erythrocyte CD59 is expressed on the cell surface only in the glycosylated form (ch.3. Fig.7) suggesting that the GPI anchor may play a role in controlling sequon occupancy.

2.3 Sialylation of oligosaccharides

The weak anion exchange chromatography profiles of the glycans associated with the leucocyte antigens studied in this thesis are shown in figure 2. The percentage of neutral structures on the four proteins, all expressed in CHO-K1 cells, ranged from 0-90%; the chimera CD5/CD4d3,4 contained only charged glycans, while at the other extreme 90% of the glycans on CD2 were neutral. The charge associated with complex sugars was invariably due to sialic acid. The CHO cell line does not express α-2,6 sialyl transferase, and, consistent with this, all the terminal sialic acid linkages associated with soluble CD59 were of the α2,3 type. In contrast, erythrocytes express both α2,3 and α2,6 sialyl transferases, and the erythrocyte CD59 N-linked oligosaccharides contained both sialic acid linkages (ch.3). All the complex sugars were present as neutral and sialylated structures although not every galactose residue was

sialylated and tri-antennary glycans were only mono or di-sialylated. The subcellular location of sialyl transferase extends from the cis to the trans Golgi (Weinstein et al 1982), suggesting that addition of sialic acid to terminal galactose may occur at any stage in the elongation process. The relationship between chain length and sialylation is summarised in table 2. Sialylation is more likely to occur in the early stages of processing (Merkle and Cummings 1987, Edge and Spiro 1985, Fukuda et al 1984) and the incidence falls off as the chains become more elongated: shorter oligosaccharides (less than 3 linear units of the repeating disaccharide Galβ1,4GlcNAcβ1) more often contain α2,6-linked sialic acid, while the α2,3 linkage is more common in the relatively less sialylated longer polylactosamine extensions (Merkle and Cummings 1987).

Glycoprotein/ cell line	% complex N-links with 1 unit of Galβ1,4GlcNAcβ1	% charged N-links	%2,3SA	% 2,6SA
CD5d1 (CHO)	100%	83%	100%	0%
CD5/CD4d3,4 (CHO)	95%	100%	100%	0%
CD2 (CHO)	79%	10%	100%	0%
sCD59 (CHO)	64%	65%	100%	0%
Hu erythrocyte CD59	50%	66%	85%	15%

Table 2

These data, which indicate that the sugar chains attached to the recombinant proteins in this study contain only α2,3-linked sialic acid, are consistent with previous studies (Ashford 1993) indicating that CHO-K1 cells lack Galα2,6-sialyl transferase. The trend in these data suggests that glycoproteins with a higher percentage of short structures (defined as 1 linear unit of the repeating disaccharide Galβ1,4GlcNAcβ1/arm) are more extensively sialylated. CD5d1 is glycosylated exclusively with glycans of this type and 83% are sialylated. The CD5/CD4d3,4 chimera which contains 95% of such glycans is 100% sialylated. However, sCD59, which contains 64% is only 65% sialylated; this result may partly be explained by the analysis of the glycans separated by charge into NA1 and NA2 pools. This indicates that only one or two branches of the triantennary glycans (approximately 10% of the structures) are sialylated, although the branches may each contain a single disaccharide unit. Clearly there are other factors involved in modulating sialylation apart from the length of the glycan chain and the

proximity of the protein matrix, for example in the case of CD2, which on this basis is anomalous. Asn 65 in CD2 may be protected from the processing enzymes by the protein structure (ch.2C, 3.7). Interestingly, it has been suggested that an additional factor may be the requirement for a recognition sequence for sialyl transferase in the protein. In support of this, sequence analysis of cloned transferases has identified a conserved peptide motif (the sialyl motif) embedded within otherwise dissimilar sequences (Kitagawa et al 1993). However recently this motif has been shown to bind the donor substrate, CMP-NeuNAc, (Datta and Paulson 1995) rather than the glycoprotein substrate.

The study of the human erythrocyte (HE) glycoprotein, CD59, (ch3) showed that approximately 50% of the biantennary N-glycans contained one Galβ1,4GlcNAcβ1 unit/arm. The remaining sugars were larger structures. 66% of the total population of N-linked sugars were sialylated; 85% of the sialylated glycans contained α2,3 linked sialic acid, while 15% (which may be all monosialylated glycans) were α2,6 linked. From the generalisations discussed above it appears probable that the large polyiactosamine structures on HECD59 are more likely to contain sialic acid in the α2,3 linkage than the smaller glycans. Interestingly, a comparison of the glycosylation of recombinant erythropoietin (CHO cells) with native erythropoietin isolated from human urine has showed that the former contains only SA α2,3 Gal linkages while the native material contains 40% of SA α2,6 linkages (Takeuchi et al 1988). These data support the finding that the CHO cell line does not contain α2,6 sialyl transferase and, in addition, indicate that the partially processed glycoforms of both CD59 and erythropoetin contain the necessary information to be recognised by both enzymes.

3. O-glycans associated with the leucocyte antigens in this study

Galβ1,3/4GalNAc O-linked glycans were identified on CD5d1 (8 amino acid 'link' region) and CD5/CD4d3,4 (23 amino acid link region). The chimera has a higher proportion of O-linked sugars than CD5d1 suggesting that the inclusion of a region rich in threonine and proline residues may be a means of introducing O-linked glycans into a recombinant protein. In contrast sCD2 and sHuCD59 expressed in CHO cells are not O-glycosylated. Interestingly, human erythrocyte (HuE) CD59 does contain O-linked glycans (ch. 3) although the glycosylation site on HuECD59 has not been determined; these data suggest that there may be a conformational change in the truncated sCD59 protein which precludes O-glycosylation, or that, in this case, O-glycosylation requires the presence of the GPI anchor.

4. Conclusion

The glycosylation of proteins is a complex event which is in part controlled by the interactions of the specific glycosylation machinery within a cell with each partially processed glycoform, and both glycan and protein are involved. The unique structure of each glycoform is defined both by the oligosaccharide and the local protein structure at the site to which it is attached. Initiation of glycosylation is controlled by the protein structure in as much as the processing of oligomannose structures requires the mannosidase enzymes to access the $Man_9GlcNAc_2$

precursor within the local protein conformation. The blocking of the α1,3 N-acetylglucosaminyl transferase II, which gives rise to hybrid glycans, is also primarily a function of the protein. Interestingly, mutant mice in which the gene which encodes GlcNAc transferase I was deleted died at mid-gestation, mainly from abnormalities in neuronal development. This means of interfering with processing has indicated an essential role for complex and/or hybrid N-linked carbohydrates (Ioffe and Stanley 1994). The proportion of polylactosamine type structures and the percentage of charged structures depends on the outcome of a competition for outer arm GlcNAc by galactosyl tranferase and sialyl transferase; the less processed the glycan chain the more likely it is to terminate with sialic acid suggesting that the affinity of the sialyl tranferase is enhanced by the proximity of the protein and shorter glycans. The type of sialic acid linkage is also related to glycan chain length and α2,6 linked sialic acid is more often associated with shorter chains than the α2,3 type.

As a result of these multiple interactions most glycoproteins consist of mixtures of glycoforms, and in the following chapters the composition and some possible functions for glycoform populations have been explored in a number of systems including human erythrocyte CD59, RNase, transferrin, IgG and tPA.

References to chapter 2:

Ashford, D.A., Alafi, C.D., Gamble, V.M., Mackay, D., Rademacher, T.W., Williams, P.J., Dwek, R.A., Barciay, A.N., Davis, S.J., Somoza, C., Ward, H.A. and Wiliiams, A.F. (1993) J. Biol. Chem. 268 3260-3267 Site-specific glycosyiation of recombinant rat and human soluble CD4 variants expressed in chinese hamster ovary ceils.

Beyers, A.D., Barclay, A.N., Law, D.A., He, Q. and Williams, A.F. (1989) immunol. Rev. 111 59-77 Activation of T-lymphocytes via monoclonal antibodies against rat ceii surface antigens with particular reference to CD2 antigens.

Beyers, A.D., Spruyt, L.L. and Williams A.F. (1992) Proc. Natl. Acad. Sci. 89 2945-2949 Molecular Associations between the T-lymphocyte antigen receptor complex and the surface antigens CD2, CD4, or CD8 and CD5.

Bodian, D.L., Jones, E.Y., Harlos, K., Stuart, D.I., and Davis, S.J. (1994) Current Biology Structure 2 755-766 Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5A resolution.

Brady, R.L., Dodson, E.J., Dodson, G.G., Lange, G., Davis, S.J., Williams, A.F. and Barclay, A.N. (1993) Science 260 979-983 Crystal Structure of domains 3 and 4 of rat CD4: reiation to NH2 terminal domain.

Brisson, J-R., and Carver, J.P. (1983b) Biochemistry 22 3680-3689 Solution conformation of asparagine-linked oiigosaccharides α(1-6)- linked moiety.

Dahms, N.M. and Hart, G.W. (1986) J. Biol. Chem. 261 13186-13196 Influence of quaternary structure on glycosylation.

Datta, A.K. and Pauison, J.C. (1995) J. Biol. Chem. 270 1497-1500 The sialyltransferase 'sialyl motif' participates in binding the donor substrate CMP-NeuAc.

Davis, A., Simmons, D.L., Hale, G., Harrison, R.A., Tighe, H., Lachman, P.J. and Waidmann, H. (1989) J. Exp. Med. 170 637-654 CD59, an Ly-6 like protein expressed in human lymphoid cells regulates the action of the complement membrane attack complex on homologous cells.

Davis, S.J., Davies, E.A., Barclay, A.N., Daenke, S., Bodian, D.L., Jones, E.Y., Stuart, D.L., Butters, T.D., Dwek, R.A. and van der Merwe, P.A. (1995) J. Biol. Chem. 270 369-375 Ligand binding by the immunoglobulin superfamily recognition molecule CD2 is glycosyiation independent.

Davis, S.J., Puklavec, M.J., Ashford, D.A., Harlos, K., Jones, Y., Stuart, D.I. and Williams, A. (1993) Protein Engineering 6 229-232 Expression of solubie recombinant glycoproteins with predefined glycosylation: appiication to the crystallisation of the T-cell glycoprotein CD2.

Deckert, M., Kubar, J., Zoccola, D., Bernard-Pomler, G., Angelisova, P., Horejsi, V. and Bernard, A. (1992) Eur. J. Immunol. 22 2943-2947 CD59 molecule: A second ligand for CD2 in T-cell adheslon.

Driscoll, P.C., Cyster, J.G., Campbell, I.D. and Williams A.F. (1991) Nature 353 762-766 Structure of domain 1 of rat T-lymphocyte CD2 antigen.

Edge, A.S.B. and Spiro, R.G. (1985) J. Biol. Chem. 260 15332-15338 Thyroid cell surface glycoproteins. Nature and disposition of carbohydrate units and evaluation of their blood group activity.

Ferguson, M.A.J. (1991) Current Opinion in Structural Biology 1 522-529 Lipid anchors on membrane proteins.

Freeman, M., Ashkenas, J., Rees, D.J.G., Kingsley, D.M., Copeland, N.G., Jenkins, N.A. and Krieger,M (1990) Proc. Nat. Acad. Sci. USA 87 8810-8814 An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type ii murine macrophage scavenger receptors.

Fukuda, M. (1991) J. Bioi. Chem. 266 21327-21342 Lysosomal membrane glycoproteins.

Fukuda, M., Dell, A., Oates, J.E. and Fukuda, M.N. (1984) J. Biol. Chem. 259 13 8260-8273 Structure of branched lactosaminoglycan, the carbohydrate moiety of band 3 isolated from human erythrocytes.

Fukuda, M. (1994) Cell surface carbohydrates: cell type specific expression - in Molecular Glycobiology eds Fukuda, M. and Hindsgaul, O. IRL Press.

Ioffe, E. and Stanley, P. (1994) Proc. Nat. Acad. Sci. USA 91 728-732 Mice lacking N-acetylglucosaminyltransferase I activity die at mid gestation, revealing an essential role for complex or hybrid N-linked carbohydrates.

Jones, E.Y., Davis, S. J., Williams, A.F., Harios, K. and Stuart, D.I. (1992) Nature 360 232-238 Crystal structure at 2.8A resolution of a soluble form of the ceii adhesion molecule CD2.

Kieffer, B., Driscoll, P.C., Campbell, I.D., Willis, A.C., van der Merwe, P.A. and Davis, S.J. (1994) Biochemistry 33 4471-4482 Three dimensional structure of the extracellular region of the complement regulatory protein CD59, a new cell surface protein domain related to snake venom neurotoxins.

Kitagawa H. and Paulson J.C. (1994) J. Biol. Chem. 269 1394 Cloning of a novel α2,3 sialyl tranferase that sialylates glycoprotein and glycolipid carbohydrate groups.

Knoppers, M.H., Reinhold, B.B. and Recny M.A. (1992) Poster T199 The Protein Society Symposium Minimum N-glycan branching structure required to maintain the conformational stability of the CD2 adhesion domain.

Lachman, P.J. (1991) Immunology Today 312-315 The control of homologous lysis.

Natsuka, S. and Lowe, J.B. (1994) Current Opinion in Structural Biology 4 683-691 Enzymes involved in mammalian biosynthesis.

McConville, M.J. and Ferguson, M.A.J. (1993) Biochem. J. 294, 305-324. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes.

Matsui, T., Mizuochi, T., Titani, K., Okinaga, T., Hoshi, M., Bousefield, G.R., Sugino, H., and Ward, D. (1994) Biochemistry 33 14039-14048. Structural analysis of N-linked oligosaccharides of equine chorionic gonadotropin and lutropin β-subunits.

Merkle, R.K. and Cummings, R.D. (1987) J. Biol. Chem. 262 8179-8189 Relationship of the terminal sequences to the length of the poly-N-acetyllactosamine chains in asparagine-linked oligosaccharides from mouse lymphoma cell line BW5147.

Metzler, M., Gertz, A., Sarkar, M., Schacter, H., Schrader, J.W. and Marth J.D. (1994) EMBO J. 13 2056-2065 Complex oligosaccharides are required for morphogenic events during post-implantation development.

Ninomiya, H., Stewart, B.H., Rollins, S.A., Zhao, J., Bothwell, A.L.M. and Sims, P.J. (1992) J. Biol. Chem. 267 8404-8410 Contribution of the N-linked Carbohydrate of Erythrocyte Antigen CD59 to its Complement Inhibitory Activity.

Parekh, R.B., Tse, A.G.C., Dwek, R.A., Williams, A.F. and Rademacher, T.W. (1987) EMBO J 6 1233-1244 Tissue-Specific N-Glycosylation, Site Specific Oligosaccharide Patterns and Lentil Lectin recognition of Rat Thy-1.

Parekh, R.B., Dwek, R.A., Rudd, P.M., Thomas, J.R., Rademacher, T.W. Warren, T., Wun, T-C., Hebert, B., Reitz, B., Palmier, M., Ramabhadran, T. and Teimeir, D.C. (1989b) Biochemistry 28 7670-7679 N-glycosylation and *in vitro* enzymatic activity of human recombinant tissue plasminogen activator expressed in chinese hamster ovary ceils and a murine cell line.

R.B. Parekh, R.A. Dwek, J.R. Thomas, G. Opdenakker, and T. Rademacher; A.W. Wittwer, S.H. Howard, R. Nelson, N. Siegel, M.G. Jennings, N. K. Harakas and J. Feder (1989) Biochemistry 28, 7644-7661 Cell-Type Specific and Site Specific N-Glycosylation of Type I and Type II Human Tissue Plasminogen Activator.

Parish, C.R., Recny, M.A., Knoppers, M.H., Waldron, J.C. and Warren H.S. (1993) J. Immunol. 150 4833-4843 Detection of a glycosylation dependent ligand for the T-lymphocyte cell adhesion molecule CD2 using a novel multimeric recombinant CD2-binding assay.

Patel, T., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A. and Parekh, R.B. (1993) Biochemistry 32 679-693 Use of hydrazine to release In intact and unreduced form both N-and O-linked oligosaccharides from glycoproteins.

Qin, L., Chavin, K.D., Lin, J., Yagita, H. and Bromberg, J.S. (1994) J.Exp. Med. 179 341-346 Anti-CD2 receptor and anti CD2 ligand (CD48) synergise to prolong allograft survival.

Recny, M.A., Luther, M.A., Knoppers, M.H., Neinhardt, E.A., Khandekar, S.S., Concino, M.F., Schimke, P.A., Francis, M.A., Moebius, U., Reinhold, B., Reinhold, V.N. and Reinherz, E.L. (1992) J. Biol. Chem. 267 31 22428-22434 N-glycosylation is required for human CD2 adhesion Functions.

Sandrin, M.S., Mouhtouris, E., Vaughan, H.A., Warren, H.S. and Parish, C.R. (1993) J. Immunology 151 4606-4613 CD2 is a low affinity ligand for human CD52.

Selvarajez,, P., Plunkett, M.L., Dustin, M.E., Sanders, S., Shaw, S. and Springer, T.A. (1987) Nature (Lond.) 326 400-402 The T-lymphocyte glycoprotein CD2 binds the cell surface ligand LFA-3.

Sherman, G.B., Wolfe, M.W., Farmerie, T.A., Clay, C.M., Threadgill, D.S., Sharp, D.C. and Nilson, J.H. 1992 Mol. Endocrinol. 6 951-959

Somoza, C., Driscoll, P.C., Cyster, J.G. and Williams, A.F. (1992) J. Exp. Med. 178 549-558 Mutational analysis of the CD2/CD58 Interaction: the binding site for CD58 lies on one face of the first domain of CD2.

Sugino, H., Bousefield, G.R., Moore, W.T., Jr., and Ward, D.N. (1987) J. Biol. Chem. 262, 8603-8609 Structural studies on equine glycoprotein hormone.

Takuechi, M., Takasaki, S., Shimada, M., and Kobata, A. (1990) J. Biol. Chem. 265 12127-12130 Role of sugar chains in the *in vitro* biological activity of erythropoeitin produced in recombinant Chinese hamster ovary cells.

Tomita, M., Tobe,T., Choi-Miura, N., Nakano, Y., Kusano, M. and Oda, E. (1991) Abstracts XIVth Int. Complement Workshop p.233 Structural analysis of MACIF (CD59).

van der Merwe, A., Barclay, N., Mason, D., Davies, E., Morgan, B.P., Tone, M., Krishnam, A.K.C., Laneiii, C., and Davis, S. (1994) Biochemistry 33 10149-10160. Human cell adhesion molecule CD2 binds CD58 (LFA3) with a low affinity and extremely fast dissociation rate, but does not bind CD48 or CD59.

van der Merwe, P.A., McNamee, P.N., Davies, E.A., Barclay, A.N. and Davis, S.J. (1995) Current Biology 5 74-84 Topology of the CD2-CD48 cell-adhesion molecule complex: implications for antigen recognition by T cells.

van de Velde, H., von Hoegen, I., Luo,W., Parnes, J.R., and Thielemans, K. (1991) Nature 351 662-665 The B-cell surface protein CD72/Lyb-2 is the ligand for CD5.

Walsh, L.A., Tone, M., Thiru, S. and Waldman, H. (1992) Tissue Antigens 40 213-220 CD59 antigen- a multifunctional molecule.

Wang, J., Yan, Y., Garrett, T.P.J., Liu, J., Rodgers, D.W., Garlick, R.L., Tarr, G.E., Husain, Y., Reinherz, E.L. and Harrison, S.C. (1990) Nature 348 411-418 Atomic structure of a fragment of CD4 containing two immunoglobulin-like domains.

Wang, W-C., Lee, N., Aoki, D., Fukuda, M.N. and Fukuda, M. (1991) J. Biol. Chem. 266 23185-23190 The poly-N-lactosamines attached to lysosomal membrane glycoproteins are increased by the prolonged association with the Goigi complex.

Williams, A.F. (1987) Immunology Today 8 298-303 A year in the life of the immunoglobulin superfamily.

Williams, A.F. and Barclay, A.N. (1988) Ann. Rev. Immunol. 6 381-405 The Immunoglobulin Superfamily - domains for cell surface recognition.

Withka, J.M., Wyss, D.F., Wagner, G., Arulandam, A.R.N., Reinhertz, E.L. and Recny, M.A. (1993) Structure 1, 69-81 Structure of the glycosylated adhesion domain of human T-lymphocyte glycoprotein CD2.

Wyss, D.F., Choi, J.S. and Wagner, G. (1995) Biochemistry 34 1622-1634 Composition and sequence specific resonance assignments of the heterogeneous N-linked glycan in the 13.6kDa adhesion domain of human CD2 as determined by NMR on the intact glycoprotein.

Yet, M-G., Shoa, M-C. and Wold, F. (1988) FASEB J. 2 22-31 Effects of the protein matrix on glycan processing in glycoproteins.

Chapter 3

The glycosylation of the human erythrocyte complement regulatory protein, CD59

Acknowledgements:
(a) Human erythrocyte CD59 was prepared by Dr. Paul Morgan (University of Wales Medical College).
(b) The glycan anchor was analysed while visiting the University of Dundee, under the supervision of Professor Michael Ferguson.
(c) The GCMS composition analysis was also carried out in Dundee in conjunction with Professor Michael Ferguson.
(d) The molecular modelling was done in collaboration with Dr. Mark Wormald, Dr. Simon Davis and Dr. Paul Driscoll.

Abbreviations: MAC: membrane Attack Complex; GPI: glycosylphosphatidylinositol; WAX: weak anion exchange chromatography, RAAM: Reagent Array Analysis Method; 2AB: 2-aminobenzidine; CD: cluster differentiation; igSF: immunoglobulin superfamily; N: neutral glycans; NA1, NA2, NA3, NA4: glycans with 1,2,3 or 4 negative charges respectively; CHO: Chinese Hamster Ovary; OGS: Oxford GlycoSystems: MALDI MS Matrix Assisted Laser Desorption Mass Spectrometry; P4 GPC: P4 Gel Permeation Chromatography; RAAM: Reagent Array Analysis Method; GCMS: gas chromatography/mass spectrometry; ANTS: sodium amino napthalene trisulphonic acid; FACE: Fluorophore Assisted Carbohydrate Electrophoresis; NDV; Newcastle Disease Virus; JBAM: Jack Bean α-mannosidase; APAM: *Aspergillus Phoenicus* (Saitoi) α1,2 mannosidase; PNGase F: Peptide N-glycosidase F; H: hexose, N; N-acetyl hexosaminidase; F: fucose; MRRF: AHM: anhydromannitol; Molar Relative Response Factor; TMS: tri-methylsilyl.

Abstract

Lysis of invading micro-organisms by the membrane attack complex (MAC) of the complement system is an integral part of the innate immune response. Complement activation involves powerful effector mechanisms against which host cells are protected by homologous restriction. One of the molecules involved in this process is CD59, a cell surface glycoprotein with a glycosylphosphatidyl inositol (GPI) anchor. CD59 is believed to bind to the complement proteins C8 and/or C9 in the activated macrophage attack complex (MAC), preventing the formation of the fully developed complex. In this study the glycosylation of the human erythrocyte CD59 glcoprotein and its glycan anchor have been examined. The principle conclusions are as follows: (a) 32% of the glycans attached to the protein were neutral; 68% contained from 1 to 3 sialic acid residues. (b) Over 50% of the N-linked glycans attached at Asn 18 were polylactosamine tri- and tetra- antennary structures larger than 17 glucose units. (c) An O-linked disaccharide, GaiGalNAc, was recovered in approximately the same molar proportions as the N-linked glycans. (d) 95% of the glycans associated with the GPI anchor were Manα1-2Manα1-6 Manα1-4GlcNH. (e) The ratio of lipid to inositol was 3:1 suggesting palmitoylation of the inositol ring. Some of the implications of this study are: (a) In common with another GPI anchored erythrocte membrane protein, acetyl cholinesterase, the conserved Man_3GlcNH core has not been further processed. This first comparison of two GPI anchored proteins from the same cell suggests that modifications to the core glycans may be cell-type specific. (b) Palmitoylation of the inositol ring reduces the possibility that CD59 wiii be cleaved from the erythrocyte cell surface by naturally occurring phospholipases. It also suggests that CD59 will not readily migrate from the erythrocyte surface and compromise homologous restriction by solubilising in membranes of non-host cells. (c) Enzyme digestions and examination of molecular models indicated that the asparagine-sugar amide linkage of CD59 is not protected from the proposed active sites of PNGase F by the local protein conformation. (d) The N-linked oligosaccharides (size range 3-6nM) attached to CD59 (diameter approximately 12nM) at Asn18CysSer are close to the membrane surface. The glycans may orient the N-terminal active site of CD59 towards the binding site on the C5b-9 complex by restricting the conformationai space availabie to the protein and limiting its interaction with the lipid bilayer.

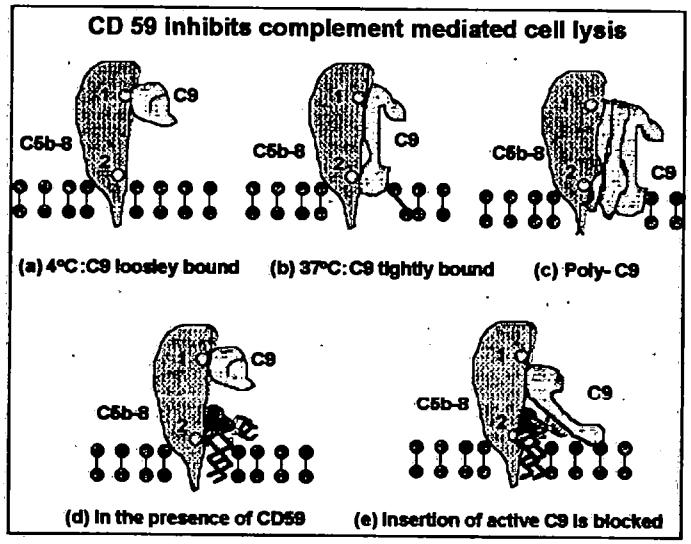


Figure 1: Schematic representation of the proposed mechanism of CD59 function

(Taken from Meri et al (1990) Immunology 71, 1-9)

Two hypothetical binding sites (1 and 2) on C5b-8 are envisaged. (a) At 4°C C9 is loosely bound at site 1. (b) At 37°C C9 unfolds and interacts with site 2. (c) Membrane damage follows exposure of hydrophobic interaction sites with the lipid bilayer and by formation of transmembrane channels. In the presence of CD59 the initial binding of C9 to C5b-8 occurs (a), but the subsequent insertion of C9 and interaction with site 2 on C5b-8 is prevented. Inhibition of the interaction between C5b-8 and C9 may be due to blocking of site 2 on C5b-8 by CD59. The association of the first C9 molecule with C5b-8-CD59 complex could be achieved by the interaction of CD59 with C9 preventing the recruitment of additional C9 molecules into the complex.

The glycosylation of the human erythrocyte complement regulatory protein, CD59

1. Background

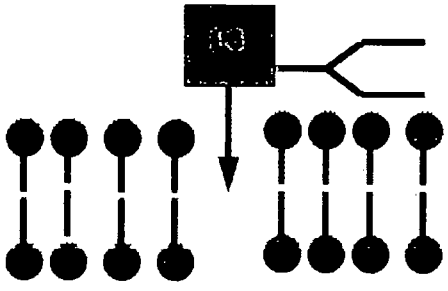
1.1 The need to control homologous lysis in erythrocytes

The earliest committed erythroid precursor cells differentiate into nucieated erythroblasts that begin to synthesise hemoglobin. Further cell differentiation and division brings about the formation of the normoblasts which, having extruded their nuclei, are released as erythrocytes through the capiilaries of the bone marrow into the blood stream. For their first two days in the bloodstream the anuclear erythrocytes (reticulocytes) contain remnants of RNA and other intracellular organelles and continue to synthesise hemoglobin. Erythrocytes have a life span of 120 days after which they are engulfed by the macrophages of the reticuloendothelial system, particularly in the liver or spleen. Some 20% of erythrocytes are prematurely destroyed in the serum, therefore protection of host ceils from further loss by homolgous lysis is an important process.

1.2 Functions and location of CD59 (also known as protectin, MIRL, HRF20, MACIF, P18, H19)

CD59 is a cell surface glycoprotein that binds to the complement proteins C8 and/or C9 in the activated MAC, protecting host cells from non-specific lysis by homologous restriction **(Fig. 1)**. Bound CD59 is believed to prevent the unfolding of the first molecule of C9 which binds to the MAC, a step that is necessary for membrane insertion (Meri et al 1990, Lachman 1991). There is then no further binding of C9 or polymerisation to form a ring (Sugita et al 1988). In addition, and controversially, it has been proposed that CD59 has a second function in mediating T-ceii adhesive interactions by synergising with CD58 via CD2. Whiie CD58 interacts with region 1 of CD2, CD59 has been reported to interact with region II (Hahn et al 1992), however this has not been confirmed at the molecular level (van der Merwe et al 1994).

CD59 is present on a wide variety of cell types, including leucocytes, erythrocytes, platelets, epithelial and endothelial cells and placental cells, and is attached to the cell surfaces via a glycophosphatidyl inositol (GPI) anchor. (Review: Waish et al 1992, Cross, 1990, Ferguson 1991) A number of soluble forms of CD59 have been found, for example in saliva, amniotic fluid, milk and urine. These may be secreted from surrounding cells or result from GPI cleavages of membrane bound molecules by endogenous phospholipases. The wide tissue distribution of CD59 appears to reflect a general need for host cell protection where the complement system is active. Increased vulnerability to complement attack has been noted where membrane bound CD59 is absent as, for example, on the erythrocytes of patients with paroxysmal nocturnal hemoglobinuria (Taguchi et al 1989). Conversely, CD59 is absent on cells which have less access to complement, such as liver parenchyma.



Human erythrocyte CD59

Figure 2: *Human erythrocyte CD59 (M.Wt.19kD) is a glycoprotein with one N-linked sugar at Asn18CysSer, a glycosylphosphatidylinositol anchor and O-linked glycan(s) at sites which have not been defined.*

CD59 normally protects only host cells from homologous complement activity (Hughes et al 1992); however, since the GPI anchor is readily soluble in most lipid bilayers, purified CD59 can be readily inserted into the membranes of many cells. For example, protection can be conferred on human PNH erythrocytes by passively incorporating HuCD59 into the cell membranes (Holguin et al 1989) while the stable transfection of human CD59 protects both CHO cells (Zhao et al 1991, Takizawa et al 1993) and rat T-cells (Walsh et al 1991) from human complement. It has been proposed that CD59 is a ligand for CD2 (Deckert et al 1992) but this has not been confirmed at the molecular level (van der Merwe et al 1994).

The number of molecules of CD59 / erythrocyte is between $2.5x10^4$ and $5x10^4$. On endothelial cells base levels are around $2x10^5$ - $3x10^5$ (Brooimans et al 1992) reflecting the fact that the endothelial lining of blood vessels is under constant complement attack. During inflammation the levels are upregulated 3-fold by activators of cellular second messenger. A number of cytokines (including IL1α, TNFα or IFNγ) have been shown to have no effect on the regulation of CD59, which appears to be controlled at the transcriptional level.

1.3 Structure of HuCD59

(i) CD59 (Fig 2) is a monomeric protein (18-20kD) containing 77 amino acids in a single domain. The precursor (Table 1) contains 128 amino acids; these include 25 amino acids of signal peptide. The carboxy-terminal segment contains a cluster of 18 mainly hydrophobic amino acids preceded by 8 polar amino acids, a sequence typical of a signal peptide for GPI anchored proteins. (ii) The cleavage site for the anchor attachment is between Asn 77 and Gly78 (Tomita et al 1991). (iii) There are two potential N-linked glycosylation sites (Asn8ProThr and Asn18CysSer), one of which (Asn 8) contains a proline residue in the sequon and has never been shown to be occupied. (iv) No O-linked sugars have been reported on any CD59 molecule analysed to date. Asialo human erythrocyte CD59 failed to bind peanut agglutinin, a lectin which is specific for one of the core sequences for O-linked sugars Galβ1,3GalNAc (Ninomiya et al 1992).

Amino acid sequence of human CD59

	Signal sequence
MGIQGGSVLF GLLLVLAVFC HSGHS	1-25
LQCYNCPN*PT ADCKTAVN**CS SDFDACLITK	1-30
AGLQVYNKCW KFEHCNFNDV TTRLRENELT	31-60
YYCCKKDLCN FNEQLEN#GGT SLSEKTVLLL	61-90
VTPFLAAA WS LHP	91-103

N*PT **Unoccupied glycosylation sequon (N8PT)**
N**CS **Glycosylation site (N18CS)**
Cleavage and anchor attachment site (Asn 77)
GGT SLSEKT - **Polar region**
VLLL VTPFLAAA WS LHP - **hydrophobic region**

Table 1. Amino acid sequence of the CD 59 precursor

Residues 1-25 are the signal sequence for entry into the endoplasmic reticulum. Residues 1-77 are the amino acids which comprise the protein, Residues 78-86 are a polar region of amino acids. Residues 87-103 constitute a hydrophobic region of amino acids which, together with the polar region, define a structure typical of a GPI anchor signal sequence.

1.4 Glycosylphosphatidyl inositol membrane anchors

GPI membrane anchors are used by a wide variety of cell surface glycoproteins (McConville et al 1993). They constitute an alternative to the hydrophobic transmembrane polypeptide domain for the anchorage of membrane bound proteins. A protein destined to receive a GPI anchor must contain an N-terminal signal sequence for entry into the lumen of the endoplasmic reticulum (Table 1) where the anchor is attached. The C-terminus of the protein must contain a GPI signal sequence, the most common feature of which is a series of 12-20 hydrophobic residues at the C-terminus of the primary translation product. This signal sequence is cleaved and replaced by a pre-assembled GPI precursor, in the case of CD59 this occurs at Asn77/Gly78 (Table 1). The C-terminal amino acid (restricted to either Cys, Asp, Asn, Gly, Arg or Ser) is linked via ethanolamine phosphate to a glycan with a conserved backbone sequence (Manα1-2Manα1-6Manα1-4GlcNH$_2$) which is in turn linked to the 6-position of the *myo*-inositol ring of phosphatidyl inositol (PI) (Fig.4a). GPI anchors contain two lipids, normally one acyl and one alkyl, attached through the phosphate on the inositol ring (Fig. 4a). In addition, a subset of anchors are palmitoylated at C3/4 on the inositol ring making them resistant to PIPLC since the action of this enzyme involves the formation of a cyclic phosphate at inositol C3 and C4 (Fig. 15). The tetrasaccharide backbone of GPI anchors may be substituted with other sugars; in common with other oligosaccharide processing events which take place in the cell, the type and extent of these modifications may be species and tissue specific and depend on the structure of the individual protein to which the anchor is attached. In contrast to the GPI anchored proteins, the more common transmembrane anchors contain a polar sequence of amino acids at the C terminus of the peptide preceded by a hydrophobic region. The polar region remains in the lumen of the Golgi, and later in the cytosol of the cell, and fixes the protein in the membrane. If the polar sequence is missing the protein may still be anchored, but with the hydrophobic region replaced with a lipid anchor. It has been suggested that GPI anchors allow greater lateral mobility of proteins through the membrane compared to those with a transmembrane domain. (Low et al 1989) In the case of CD59 this may allow rapid movement towards sites of complement attack, before lytic plugs can form.

2. Introduction to chapter 3

The N- and O-linked glycans covalently attached to human erythrocyte CD59 were released with hydrazine and partially characterised. The structures were compared with those associated with human soluble CD59, expressed in CHO cells without a glycan anchor, which were analysed in ch.2. The accessibility of the sugar-asparagine linkage was probed

with PNGase F. The glycan component of the GPI anchor was also isolated and characterised, and compared with the structures present in the glycan anchor attached to another human erythrocyte GPI anchored glycoprotein, acetyl cholinesterase. The oligosaccharides were analysed using a combination of P4 gel filtration and exoglycosidase digestions and GCMS composition analysis. Molecular modelling was used to explore a possible role for the glycans attached to CD59 in the function of the molecule in protecting erythrocytes from homologous lysis.

3. Methods

3.1 De-N-glycosylation of CD59 with PNGase F.

CD59 was incubated for 16h at 37^{o}C with *Flavobacterium meningoseptum* peptide-N-glycosidase F (analytical grade) in 20mM sodium phosphate buffer pH 7.5 containing 50mM EDTA and 0.02% azide buffer to deglycosylate the protein prior to ESMS.

3.2 Analysis of glycan moiety of the GPI anchor

3.2.1 Nitrous acid deamination and sodium borohydride reduction

The sample was dried in an Eppendorf tube in the Speed-Vac, and redissolved in 10µl (1 volume) pH 4.0 buffer (0.1 M sodium acetate adjusted to pH 4 with acetic acid). 10µl (1 volume) freshly prepared 0.5M $NaNO_2$ were added and incubated for 2.5h at room temperature. 5µl (0.5 volume) 0.8M boric acid were added. The amount of NaOH required to adjust the pH to about 10.5 was determined using buffer controls (typically equivalent to 8-10µl i.e. 0.8 - 1 volume per sample digest). This amount was added to the reaction mixture and mixed. 5µl of 36mM NaB^3H_4 (10-15 Ci/mmol) dissolved in 0.1M NaOH were added immediately (hot reduction), followed by freshly prepared 1M $NaBD_4$ in water to give a final concentration of 0.2 M $NaBD_4$ (cold reduction). The reduction mixture was acidified with 1M acetic acid. The reaction was carried out in the fume hood as tritium gas was evolved. The sample was dialysed against water using a Spectropore membrane with a 2000 daltons cut off, freeze dried and 20% dissolved in 50µl water.

3.2.2 Aqueous HF dephosphorylation

The sample was placed in a screw top Eppendorf tube, dried in a Speed-Vac, and placed on an ice/water mixture in a Dewar flask. 50µl 48% aqueous HF (BDH Aristar) at -20^{o}C were accurately measured into the tube containing the sample and into 5 blank tubes. All the tubes were incubated for 48-60 hours at 0^{o}C. The blank tubes were spun briefly in the microfuge. 250µl of saturated LiOH (prepared by shaking 10g of LiOH with 30ml of water in plastic tube) were added to one blank. After centrifuging (Microfuge 20sec) 1ml aliquots checked for pH. If pH was <3 10µl saturated LiOH were added, vortexed and the pH re-checked. The previous step was repeated until pH >3; the volume of LiOH required to give final pH of 3-5 was estimated and the reproducibility checked against other blanks. This

volume of LiOH was frozen in a labelled Eppendorf tube on dry ice. The HF digest was transferred to the frozen LiOH, vortexed, spun and transferred to original tube using the same pipette tip. The pellet was washed twice with 50µl water and supernatants transferred to original tube.

3.2.3 Re-N-acetylation

Re-N-acetylation was carried out by adding 10mg $NaHCO_3$ to saturation (approx. 100mg/ml; 1M) to the combined supernatants from 3.1.2, followed by 10µl acetic anhydride. The mixture was incubated for 10 minutes at 4°C, twice. 10µl acetic anhydride were added and the mixture incubated at room temperature for 30 minutes, desalted by passing through 0.2ml AG50X12(H^+), eluted with 1ml water. The eluate contained the glycan core (Gc), starting from the deaminated and reduced glycopeptide (dAR-gp). The sample was dried by rotary evaporation, and flash evaporated with toluene to remove residual acetic acid. The sample was dissolved in 25µl water and applied to a 3x40cm paper strip (Whatman 3MM) 10 cm from the end. Descending paper chromatography for 60h in 1-butanol:ethanol:water (4:1:0.6) was used to remove radiolabelled contaminants (labelled GPI glycans remain at the origin if phosphorylated and within 3 cm of origin if neutral). The paper was dried and scanned and the relevant region cut from the chromatogram and eluted with water (50µl/cm^2) by incubating for 5 mins and centrifuging through a syringe hung inside a rotary evaporator tube (5min at 2-3000rpm). This elution procedure was repeated 3 times using 25µl/cm^2 water and the combined eluates evaporated to dryness. The eluted glycans were passed through a mixed-bed column of 0.1ml of Chelex 1000 (Na+), over AG50X12 (H^+), over AG3X4 (OH^-), over QAE-Sephadex A25 (OH^-), and filtered through a 0.2mm syringe microfilter.

3.2.4 Dionex HPLC

This technique allows glycans to be separated on the basis of net charge at high pH.

The system was fitted with a Dionex CarboPac column which is a strong anion exchange column designed to separate carbohydrates at high pH. At high pH sugar hydroxyl groups are ionised and glycans can be separated by ion exchange using an acetate gradient. (Townsend et al Y.C. (1989). In Methods in Enzymology Vol. 179 (ed. V. Ginsburg), p 779. Academic Press, New York.) A calibration standard consisting of a glucose oligomer standard mixture (5ml of 50 mg/ml standard mix) was resolved and detected by high sensitivity pulsed amperometric detection (PAD). The elution positions of the neutral glycans are defined in Dionex Units (Du) by linear interpolation of their elution positions between two adjacent internal standards. These Du have no specific meaning but are a reliable chromatographic property which correlate with particular glycan structures. The sample was detected on line by a Raytest Romana radioactive monitor, and by hand counting using a Beckman Scintillation counter, and correlated with the standards.

3.2.5 Dionex HPLC conditions:

Buffer 1, 0.15M NaOH; buffer 2, 0.15M NaOH containing 0.25M sodium acetate. Flow rate 0.6ml/min.

Time	% buffer 1	% buffer 2
0	95	5
50	80	20
51	0	100 (wash cycle)
60	0	100 (wash cycle)
61	95	5 (re-equilibration for > 10 min)

3.2.6 Biogel P-4 gel filtration

This technique allows the separation of neutral glycans on the basis of hydrodynamic volume, giving a direct measurement of size. it can be used preparatively, without internal standards, or analytically by co-chromatographing a set of glucose oligomer internal standards with radiolabelled neutral glycans. The detection of the oligomer standards by refractive index (RI) and the neutral glycans by an on-line radioactive monitor allows accurate determination of the size of the glycans in terms of effective 'glucose units' (gu). A combination of specific exoglycosidase digestions and re-chromatography of the digests allows an analysis of the glycan structures.

The analyses described were performed using the Oxford GlycoSystems Glycomap Instrument which is based on the P4 chromatography system.

3.2.7 Exoglycosidase sequencing

1. *Aspergillus phoenicus* α1-2mannosidase (APAM) (also known as *A.Saitoi* α1-2 mannosidase.)

This enzyme is specific for α-D-Man1-2D-Man glycosidic bonds. Digestions were performed in 10μl 0.1M sodium acetate pH 5.0 containing 0.01mU (0.04mg) of APAM (Oxford GlycoSystems Ltd. X-5009) for 16-24 hours at 37°C under a toluene atmosphere at a substrate concentration of betwen 15mM and 2mM. The products were de-salted by passing through 0.2ml AG50X12 (H^+) and elution with 1ml water. The samples were evaporated to dryness and residual acetic acid removed by flash drying.

2. Jack bean α-mannosidase (JBAM)

This enzyme has a broad specificity for any terminal, non-substituted, α-D-Man residue.

Digestions were performed in 30ml enzyme (25U/ml in 0.1M sodium acetate pH5). Substrate concentration was between 10mM and 5mM, incubation was for 2-4 hours at room temperature followed by 16-24 hours at 37°C. The products were desalted with AG50X12(H^+) (0.2ml), filtered through a 0.2mm membrane and evaporated to dryness. Acetic acid was removed by flash drying with toluene.

3.3 Mass Spectrometric analysis of CD59

3.3.1 GC-MS Composition analysis of lipids, protein and total sugar

All spectra were obtained with a Hewlett-Packard 5890-5970 GC-MS system. Mass spectra were recorded using electron impact (70eV) ionisation at a source temperature of 150^{o}C and pressure of 2-4 x 10-5 Torr. The GC stage was equipped with an on-column injector. The column head pressure was 5 psi giving a He carrier gas flow rate of about 0.5 ml/min through a 30m x 0.25mm SP2380 bonded phase column (Supelco) or Econocap SE-54 bonded phase column (Alltech). The GC to MS transfer line temperature was maintained at 10^{o}C above the maximum oven temperature. The transfer line was made from a 1m x 0.25mm piece of deactivated fused silica (Supelco). The column to transfer line connections are made through glass quick seal connectors (Chromopack 4787).

3.3.2 Programme for Methanolysis/TMS sugars and lipids:

Column: SE-54; initial temperature: 140^{o}C; hold time: 2 min; gradient: 15^{o}C/min; intermediate temperature: 260^{o}C (10min); gradient 2:15^{o}C/min; final temperature: 300^{o}C (20min); mass spec. acquisition: Linear scanning (m/z 40-800).

3.4 Composition analysis of total sugars associated with CD59

The reaction vessels were prepared from heat cleaned (3h at 500^{o}C), glass capillary tubes 9SMi size J, yellow band pipette capillaries, Sigma C6148) by flame sealing one end to form a microtube. *scyllo*-Inositol (sI) (0.1mM, 4nmoles, 40µl) and water (50µl) were added to CD59 (50µg /10µl, 2nmoles) (Solution A). 80µl solution A were evaporated to dryness and then incubated overnight at 85^{o}C with 0.5M HCl/MeOH in a sealed capillary. 5ml of stock solutions containing 0.5nmoles of each component of two standards: Xyl/Fuc/Gal/Glc/sI/GlcNAc/SA and Man/Gal/GlcNAc/sI/SA and a blank were treated in the same way. After cooling, the tubes were opened, 10µl of pyridine were added to neutralise the HCl and the contents dried in the Speedvac; 50µl water were added to each. The samples were dried again to remove residual HCl followed by 20ml methanol to dehydrate the residues. 15µl TMS were added, the tubes sealed with Teflon tape and left for at least 30 mins. 1µl was injected onto the GC. The monosaccharide components were obtained by integration of the total ion chromatogram peaks as follows:

(i) the peaks of the methanolysed sugar standards were integrated to give a molar relative response factor (MRRF) for each monosaccharide versus the *scyllo*-inositol internal standard:

$MRRF_{monosacc}$ = (area monosacc. peak)/(area *scyllo*-inositol peak)

(ii) The amount of each monosaccharide (M) was calculated from the sample chromatograms as follows:

Amount of M = (area sample peak) x (amount internal standard)/(area *scyllo*-inositol internal standard peak) x ($MRRF_{monosacc}$)nmol

3.5 Analysis of N- and O- linked glycans

3.5.1 Release and re-N-acetylation of glycans

0.6mg (31.6 nmoles) CD59 was dialysed into 0.1%TFA and lyophilised. Glycans were released from CD59 by hydrazine at 95^{o}C using the GlycoPrep 1000 (Oxford GlycoSystems Ltd) optimised for maximum recovery (approximately 85%) of both N- and O-linked sugars. The recovery is assessed by de-sialylating and de-galactosylating biantennary complex sugars and then incubating with ^{3}H galactosyl transferase. The free glycans associated with the anchor were not released under these conditions.

3.5.2 ^{3}H labelling of reducing terminus

50% (estimated approximately 13.4 nmoles) of the released and re-N-acetylated sugar library was reduced with $NaB^{3}H_{4}$

3.5.3 Fluorescent labelling of reducing terminus with 2-aminobenzamide (2AB)

10% (2.68nmoles) of the free glycans were evaporated to dryness using a Savant SpeedVac. 150µl of 2-aminobenzamide were added to DMSO (350µl) and 200 µl of this solution was added to a dye. 100ml of the mixture were added to the reductant and vortexed. 5µl of the final solution were added to each glycan sample, capped and incubated at 65^{o}C for 2 hours. After the incubation the samples were centrifuged briefly and spotted onto 3MM paper in a singie transfer. A chromatography tank was saturated with butanol:ethanol:water (4:1:1) and the excess reagents eluted from the labelled glycans by ascending paper chromatography. When the dye reached the top of the paper the strip was thoroughly dried and scanned using a fluorescent instant imager. The 2AB labelled glycans remaining at the origin were eluted with water.

40% (10.7 nmoles) of the unlabelled glycans was de-sialylated and purified through mixed bed resins for mass spectrometric analysis

3.5.4 Ion exchange separation of anionic sugars

A weak anion exchange column (Vydac 301VHP575) (7.5cmx50mm) was equilibrated in water. Labelled glycans were injected in aqueous solution. Neutral structures were eluted in the void volume and charged glycans which bound to the column were successively eluted with an increasing salt gradient of 0-500mM ammonium formate pH9. The classes of charged glycans were assigned by comparison with a standard fetuin glycan library.

Gradient parameters:

Time	Flow	Curve	%A	%B
Start	1ml/min	6	0	100
12min	1ml/min	6	5	95
25min	1ml/min	6	21	79
50min	1ml/min	6	80	20
55min	1ml/min	6	100	0
65min	1ml/min	6	100	0
66min	1ml/min	6	0	100
70min	2ml/min	6	0	100
89min	2ml/min	6	0	100
90min	1ml/min	6	0	100

3.5.5 Enzyme digestions

Glycans were evaporated to dryness on a SpeedVac. 20µl of standardised enzyme solutions were added and incubated for 16h at 37°C.

(a) *Bacteriodes fragilis* endo-β–galactosidase (Oxford GlycoSystems): 0.5U/ml in 50mM sodium actetate buffer pH5.8; substrate concentration 40µM. Cleaves R-GlcNAcβ1-3Galβ1-4GlcNAc/Glc.

(b) Bovine testes galactosidase (Oxford GlycoSystems): 1-2 U/ml in 100mM citrate/phosphate buffer pH4; substrate concentration 40-100µg/ml. Cleaves Galβ1-3,4>6 GlcNAc/R.

(c) *Arthrobacter ureafaciens* neuraminidase (Oxford GlycoSystems): 1-2U/ml in 100mM sodium actetate buffer pH5; substrate concentration 5-30µM. Cleaves NeuNAcα2-6>3,8 R.

(d) Newcastle disease virus neuraminidase (Oxford GlycoSystems): 0.2U/ml in 50mM sodium actetate buffer pH5.5; substrate concentration >5µM. Cleaves NeuNAcα2-3,8 R.

3.5.6 Purification of enzyme digests prior to P4 analysis

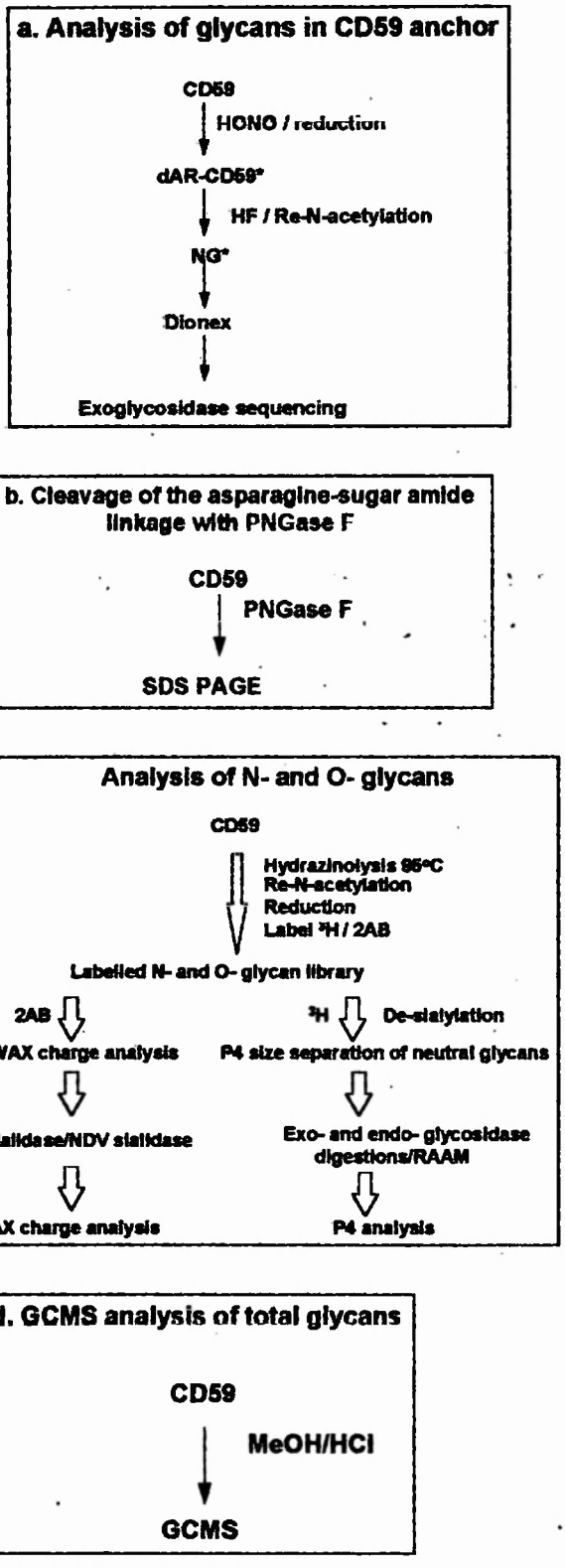
Samples were purified from protein and salts by passing through mixed bed resins of Chelex100 (Na^+)/Dowex AG50X12 (H^+)/Ag3X4A (OH^-)/QAE Sephadex A-25, as described in 3.2.3.

3.6 Molecular modelling

CD59 co-ordinates were taken from the NMR solution structure (Kieffer et al 1994).

PNGase F co-ordinates were taken from the crystal structure (Norris et al 1994).

Figure 3: Strategy for the analysis of the glycans associated with human erythrocyte CD59



4. Results

Figure 3 shows the overall scheme for the glycan analysis of the GPI anchored human erythrocyte cell surface antigen, CD59. Figure 4a-c shows the main chemical reactions involved. The strategy is in four parts: (a) analysis of the glycan part of the CD59 GPI anchor (b) examination of the Asp18CysSer-sugar amide linkage with PNGaseF (c) characterisation of the N- and O- linked glycans associated with CD59 (d) the composition analysis of all glycans attached to CD59.

4. 1 Analysis of the glycans associated with the CD59 GPI. (Fig.3a)

4.1.1 Isolation of the free neutral anchor glycans (NG)

Intact CD59 (Fig. 4a) was dialysed into water and freeze dried. The glycopeptide was released from the inositol phosphate - lipid fragment (Fig. 4b) by de-amination and reduction with nitrous acid and sodium borohydride. Sugar phosphate bonds were cleaved with aqueous HF (Fig. 4b) to yield the free glycans (NG) (Fig. 4c) which were re-N-acetylated with acetic anhydride.

4.1.2 Separation of the free anchor glycans by High Perfomance Anion Exchange Chromatography (HPAEC)

The free glycans were separated on the basis of charge at high pH by Dionex anion exchange chromatography (Fig. 5). Three glycans were resolved at 2.51 (90%), 3.07 (9%) and 3.72 (1%) Dionex units (Fig 5).The major species (90%) had a molecular volume on P4 of 4.2 gu (Fig. 6a). The elution positions of this structure on P4 and Dionex are consistent with those of Man_3AHM.

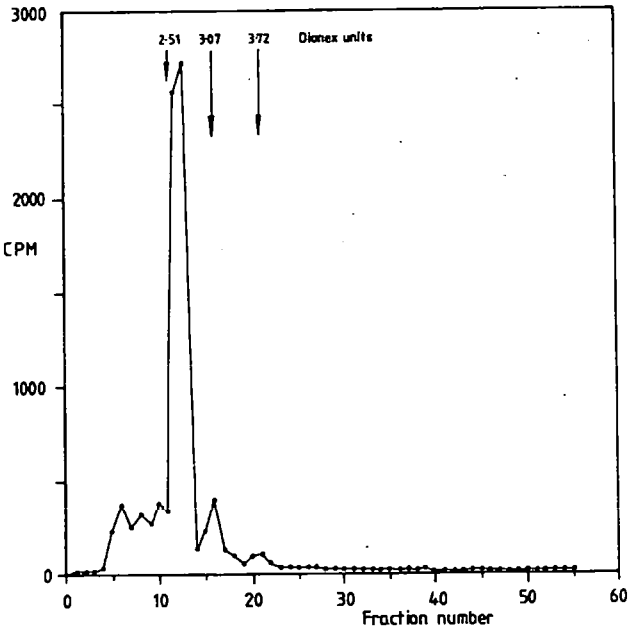


Figure 5: HPAEC (Dionex) of neutral glycans from CD59. *90% of the glycans eluted at 2.51 Dionex Units consistent with the presence of the Man_3 glycan in the anchor. 9% of the*

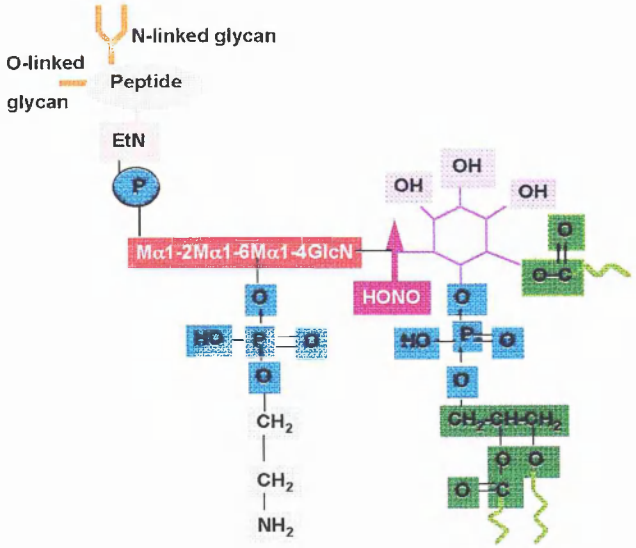


Figure 4a: Typical GPI anchor showing nitrous acid cleavage site

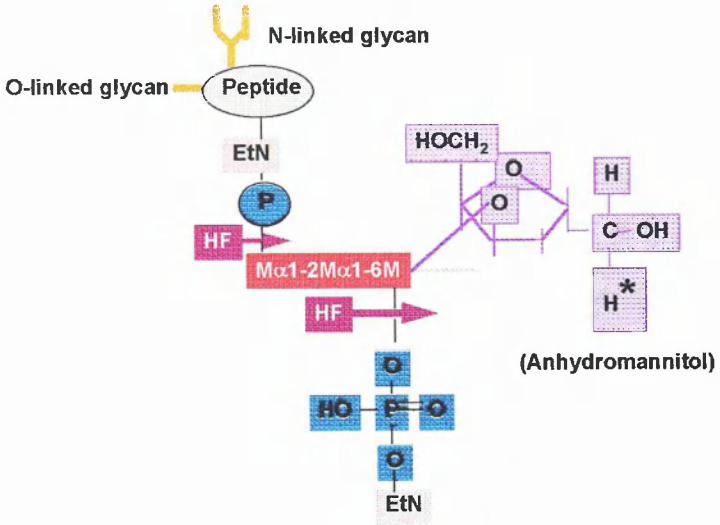


Figure 4b: De-aminated and reduced glycopeptide showing cleavage sites for dephosphorylation with HF

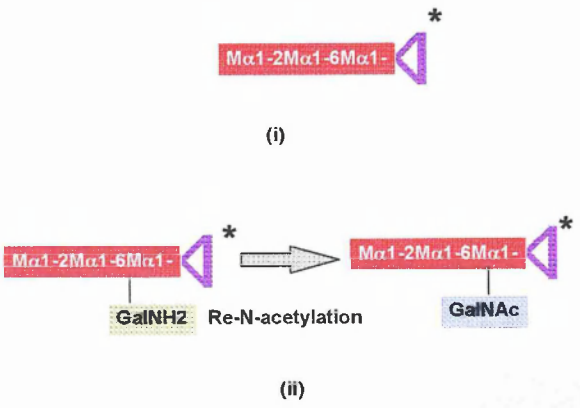


Figure 4c: Neutral glycans - products of dephosphorylation with HF

glycans eluted at 3.07 Dionex Units consistent with the presence of the Man_3 glycan substituted with GalNAc. 1% of the glycans eluted at 3.72 Dionex Units. The composition of the glycans in the 2.51 and 3.07 DU peaks was confirmed by P4 GPC.

4.1.3 Characterisation of anchor glycans by P4 GPC and exoglycosidase sequencing

The minor species (9%) had a molecular volume on P4 of 5.62 gu (Fig. 6d). The elution positions of this structure on P4 and Dionex are consistent with those of $GalNAcMan_3$ anhydromannitol (AHM). The minor species (1%) was not analysed.

The proposed structure of the major glycans (90% of total population) was confirmed by exoglycosidase sequencing and P-4 gel filtration. The 4.2gu structure was reduced with *A. Saitoi* α1-2 mannosidase to 3.2gu (Fig. 6b). This structure (Manα1,6AHM) was reduced to AHM (1.7gu) (Fig. 6c) after incubation with Jack Bean α-mannosidase which cleaves Manα1,6 linkages (in addition to Manα1,2 and Manα1,4).

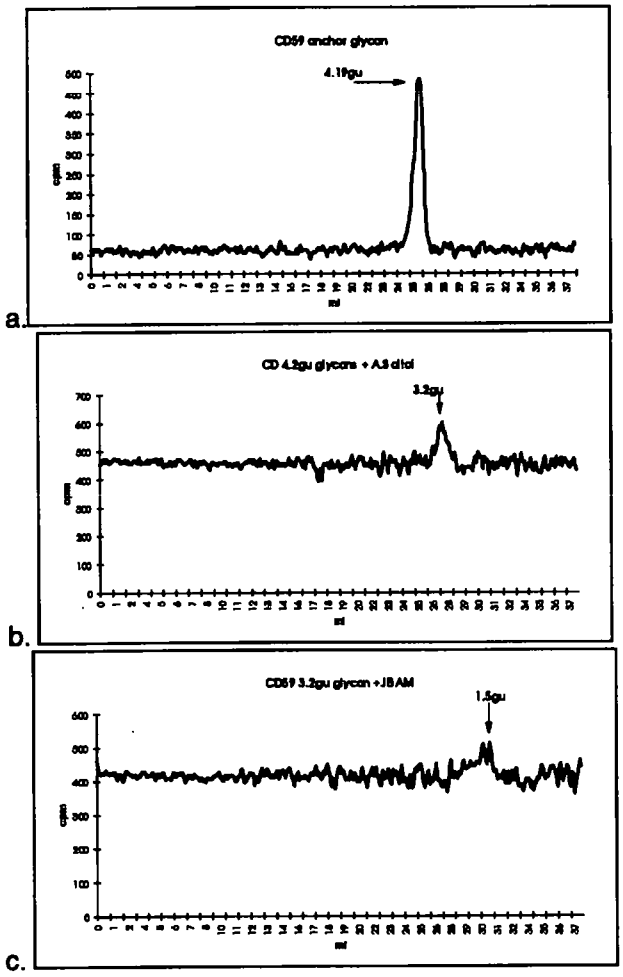


Figure 6a-c: Analysis of CD59 anchor glycans

***(a)** P4 gel permeation chromatogram of the major (90%) subset of the glycans in the CD59 anchor eluting at 4.2gu. **(b)** Digestion of glycans in 6a with α1,2 A. phoenicus mannosidase (also known as A. Saitoi) results in removal of one mannose residue and a corresponding decrease in size to 3.2gu. **(c)** Incubation of glycans from 6b with Jack bean α mannosidase*

(which cleaves Manα1,2,4 and 6 linkages) removes 2 mannose residues and reduces the glycans to 1.5gu.

d.

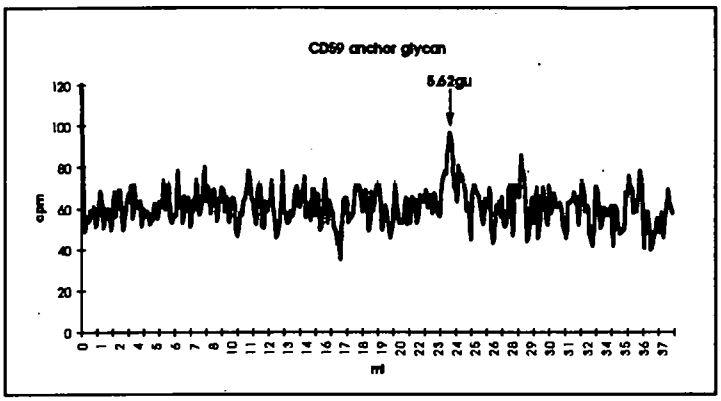


Figure 6(d): *P4 gel permeation chromatogram of the minor (9%) subset of the glycans in the CD59 anchor eluting at 5.62gu.*

4.1.4 Conclusion

The populations of CD59 anchor glycans are:

Manα1-2Manα1-6Manα1-4GlcNH 91%

Manα1-2Manα1-6Manα1-4[GalNAcβ1,4]GlcNH 9%

Insufficient material was available to determine the linkage of GalNAc to glucosamine. The assignment of a β1,4 linkage is by analogy with other GPI anchor structures and from the Dionex elution position.

4.2 The N-glycosidic linkage of CD59 is accessible to PNGase F: (Figure 3b)

Figure 7 shows a 12% SDS PAGE reducing gel of CD59 before (lane 2) and after (lane 3) digestion with PNGase F, which cleaves N-linked sugars from proteins at the N-glycosidic linkage. In lane 2 the glycoprotein can be seen as a broad band with a molecular weight range of approximately 25-19kD, consistent with the presence of a number of glycoforms. There is no evidence of any unglycosylated protein which would suggest variable site occupancy. Lane 3 indicates that approximately 80% of the sugars have been cleaved non-selectively from the protein. The de-N-glycosylated protein migrates to 15kD and is visible as a narrow band indicating that all the glycoforms have been reduced to a single structure and that the asparagine18-sugar amide linkage is accessible to PNGase F. This is consistent with the molecular modelling studies described in section 4.5.

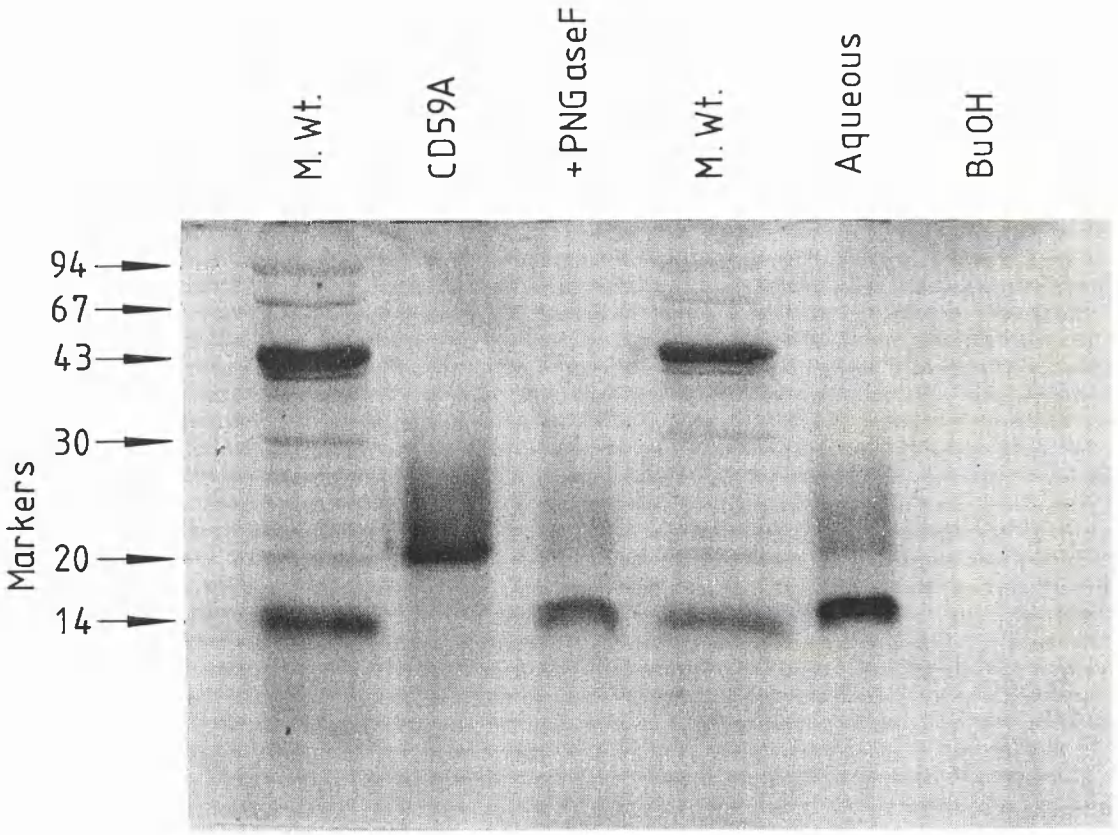


Figure 7: SDS PAGE analysis of CD59. *Tracks 1 and 4: Molecular weight standards. Tracks 2 and 3: CD59 -/+ PNGase F respectively.*

4.3 Analysis of CD59 N- and O-linked glycans: (Figure 3c)

The N- and O- glycans were released by hydrazine at 95^{o}C. 20% of the pool was labelled at the reducing terminus with 2AB and the remainder with $NaB^{3}H_{4}$.

4.3.1 Analysis by charge

The 2AB labelled glycans were analysed according to charge by weak anion exchange (WAX) chromatography (Fig. 8a). Peaks were assigned by comparison with standards of 2AB labelled fetuin glycans (Fig.8b,c). All glycans were reduced to neutral sugars following digestion with *Arthrobacter ureafaciens* neuraminidase (Fig.8d) which removes terminal α2,3- and α2,6- linked NeuNAc. All the charge could therefore be ascribed to sialic acid. After extensive digestion with Newcastle disease virus neuraminidase (which removes only α2,3 linked NeuNAc) the proportions were 76.6%:19.4%:1.38%:1.12%:1.11% (Fig.8e). 100% of the A1 glycans remained intact while over 90% of the other structures were de-sialylated. This suggests that while the OA, NA2 and NA3 glycans contain α2,3 sialic acid the A1 sugars are terminated with α2,6 linked sialic acid.

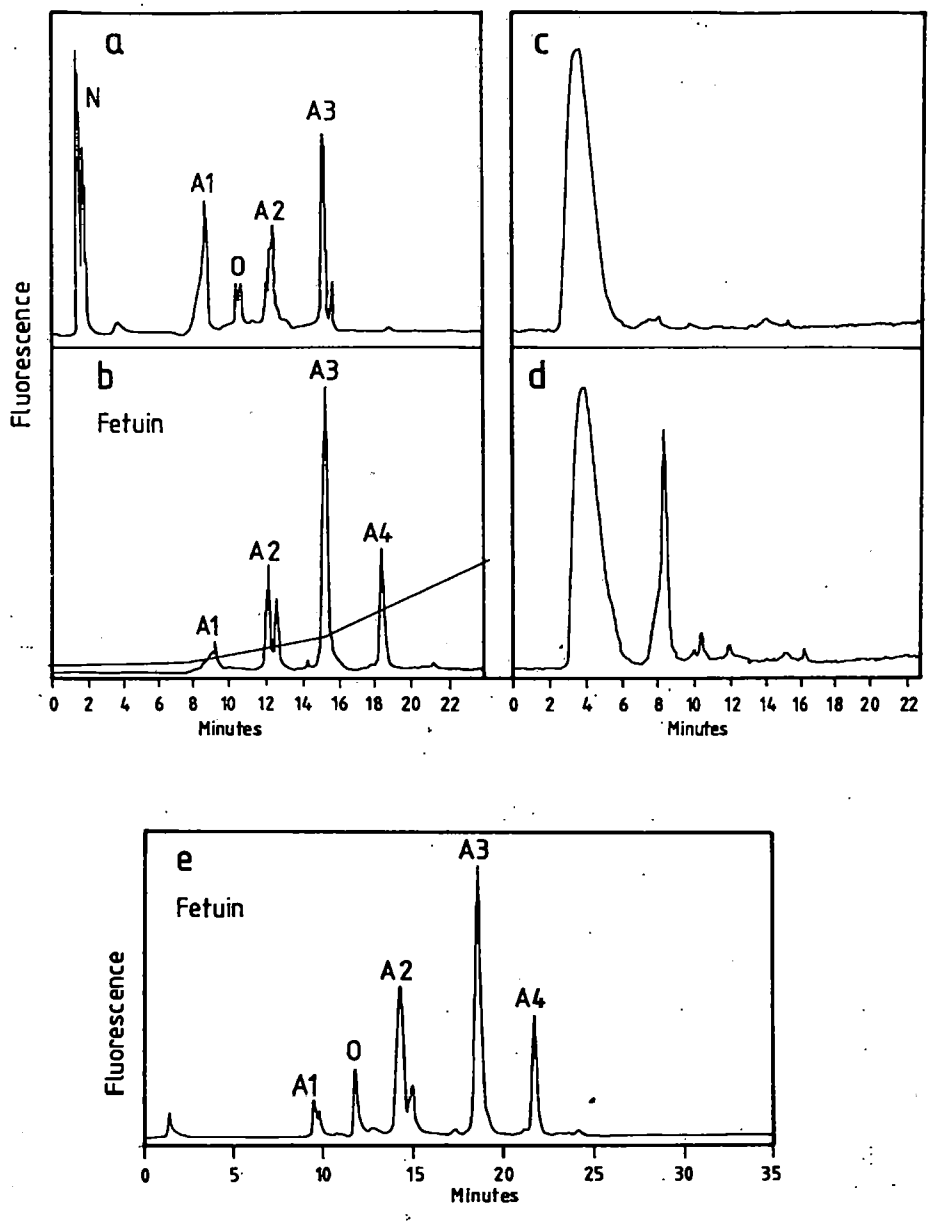


Figure 8: Analysis of CD59 glycans by mass:charge. Weak anion exchange chromatography (Vydac 301VHP575) of glycan library.

a. Resolution of 2AB labelled CD59 glycans released by hydrazinolysis. N-linked glycans carrying 1, 2, 3 or 4 anionic charges are denoted by NA1, NA2, NA3, NA4 respectively; charged O-linked glycans by OA. b. Elution positions of 2AB labelled standard fetuin N linked sugars used to assign peaks in 8a. c. Elution positions of fetuin N- and O-glycans confirming O-links elute between NA1 and NA2. d. Analysis of 2AB labelled CD59 glycans post digestion with Arthrobacter Ureafaciens sialidase. e. Analysis of 2AB labelled CD59 glycans post digestion with Newcastle Disease Virus sialidase.

Conclusion: Sialic acid content of CD59 N- and O-linked glycans is as follows:

neutral	NA1	OA	NA2	NA3
31.9%	22.2%	5.2%	17.4%	21.8%.

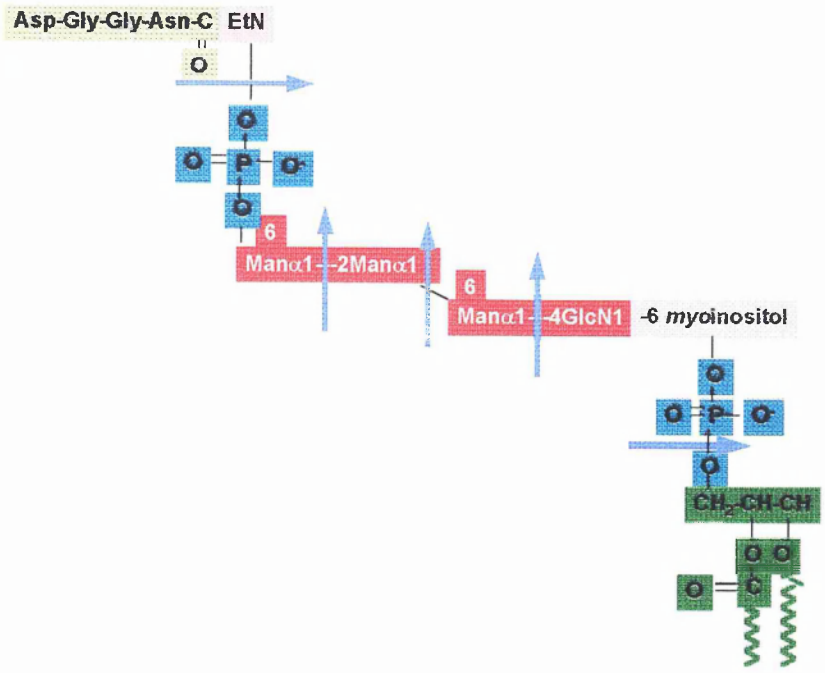


Figure 11a: Typical cleavages in acid methanolysis

4.3.2 Analysis of neutral glycans by size

The neutral glycans resulting from the neuraminidase digestion were resolved by P4 gel permeation chromatography (Fig. 9a).

Analysis of N-linked glycans

Glycans eluting between 18 and 27 gu were digested with endo β-galactosidase. This enzyme, which cleaves β-linked galactose residues within the glycan chain but not terminal galactose, gave a range of products with molecular volumes consistent with the presence of multiantennary polylactosamine structures (Fig. 9b). The 11.5 digestion product (H3N5) indicates that there are biantennary complex sugars with lactosamine extensions. The 13.5 structure (H3N6) indicates the presence of triantennary glycans with lactosamine extensions while the 18-25 gu peak suggests that the endogalactosidase digestion may not have reached completion. The 15gu fractions was identified as H5N4F using the reagent array sequencing method. The 13.8 and 11.2 structures remain to be analysed when more material is available.

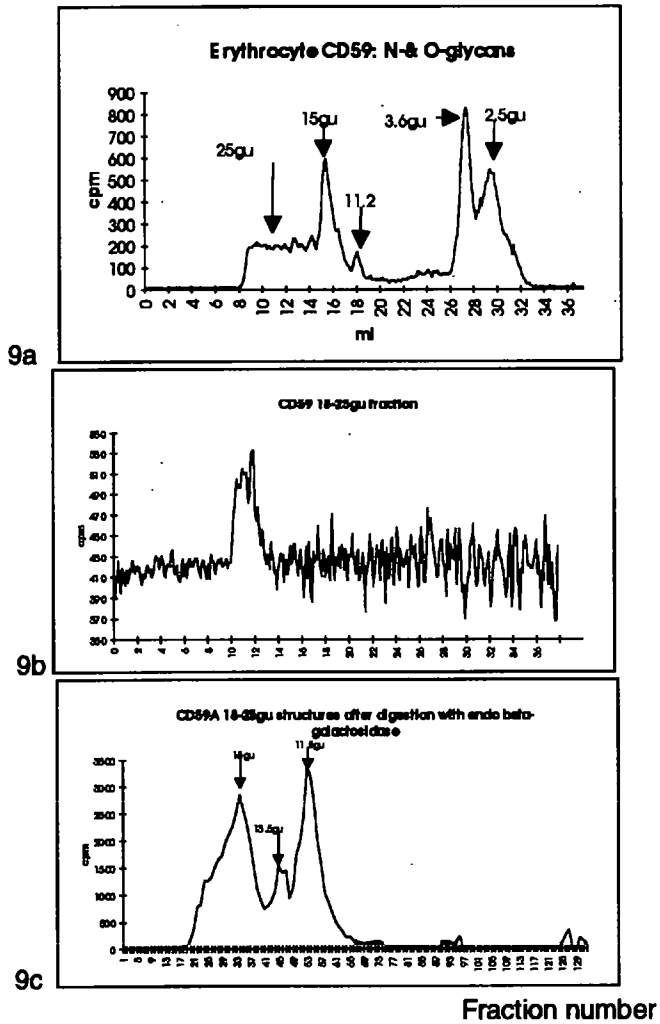


Figure 9: Analysis of neutral CD59 glycans by hydrodynamic volume using P4 gel permeation chromatography. *a. Total glycan library released by hydrazinolysis and labelled with tritium. b. Analysis of pool of 18-25 gu glycans. c. 18-25gu glycans from 9a/b after incubation with endo β-mannosidase.*

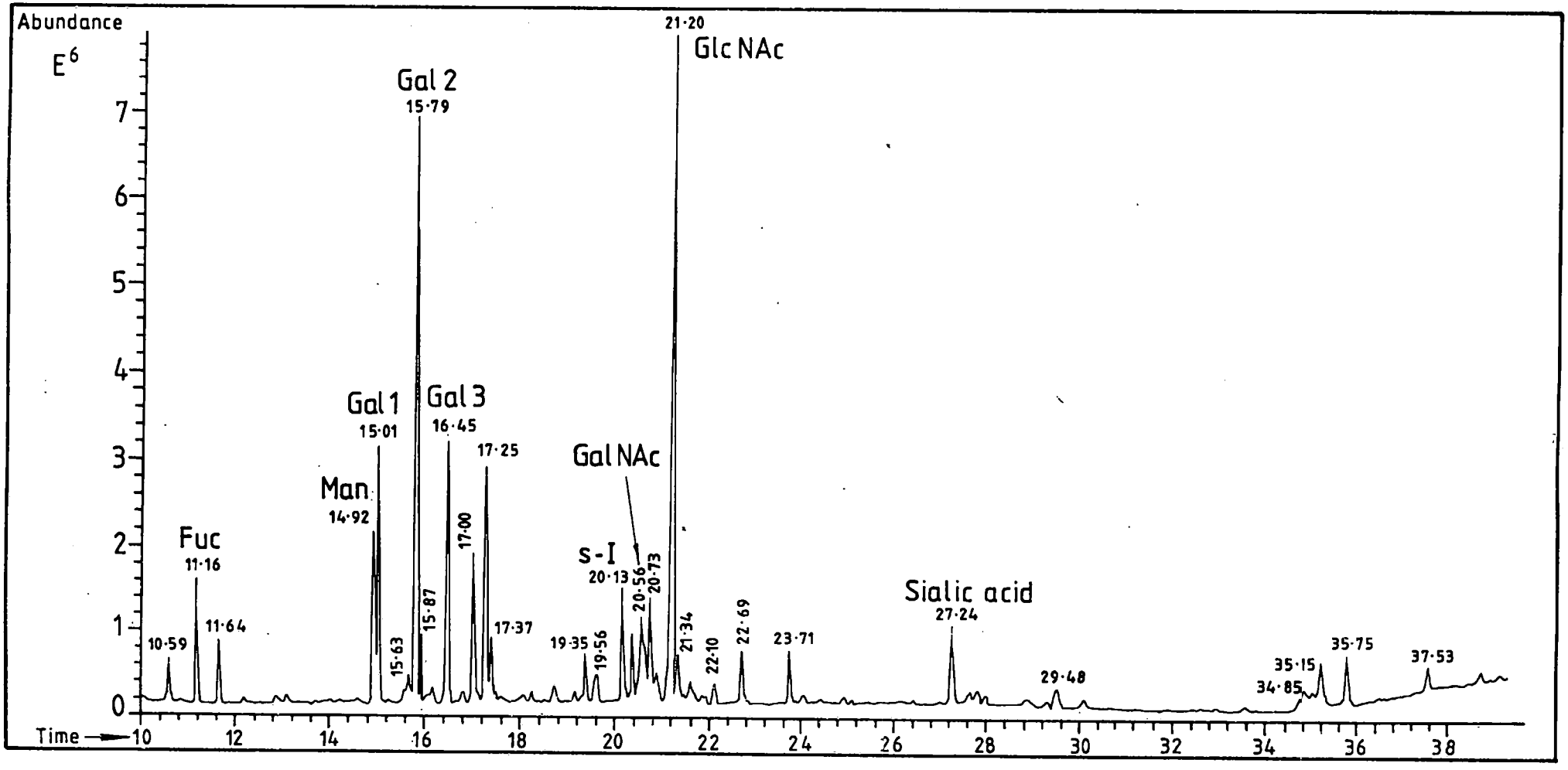


Figure 11b: GCMS composition analysis of glycans associated with CD59

Section of chromatogram showing oligosaccharides

4.3.3 Analysis of CD59 O-linked glycans by sequential exoglycosidase digestion and P4

The glycans eluting at 3.5gu were digested to 2.5gu with bovine testes β-galactosidase (Fig. 10a) The reducing terminal glycan was shown to be ^{3}HGalNAc by gas chromatography (GC) using a radioactive detector (Fig.10b). These data are consistent with the structure Galβ3,4, or 6-GalNAc. The presence of previously unreported O-linked sugars was confirmed by GCMS composition analysis (fig. 11b,c) in which the GalNAc residue, present only in O-linked glycans, was identified. The 2.5gu peak (fig. 10a) was also analysed by GC and did not contain carbohydrate.

Conclusion: 50% of the CD59 N-linked glycans are of the multiantennary polylactosamine type.

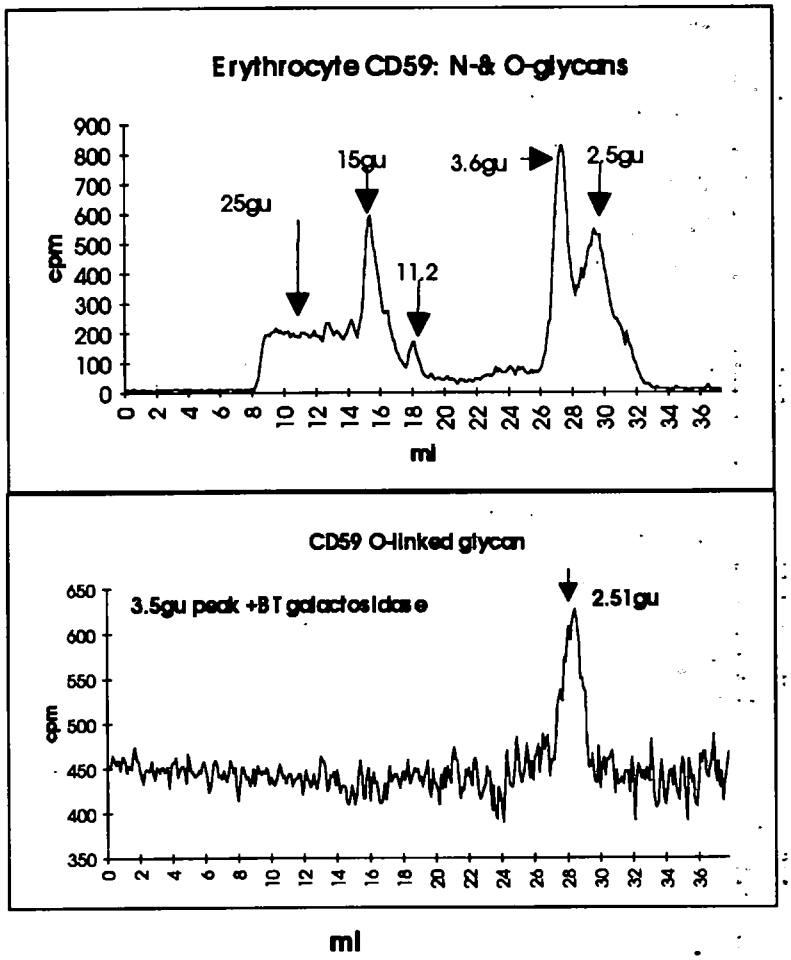


Figure 10a: Exoglycosidase analysis of the 3.6gu peak from figure 9a.

P4 gel permeation chromatogram of 3.6 gu peak following incubation with Bovine testes galactosidase shows reduction to 2.5gu consistent with the structure Gal β1-3/4GalNAc. The 2.5 gu peak in figure 9a, was analysed by radioGC and does not contain carbohydrate.

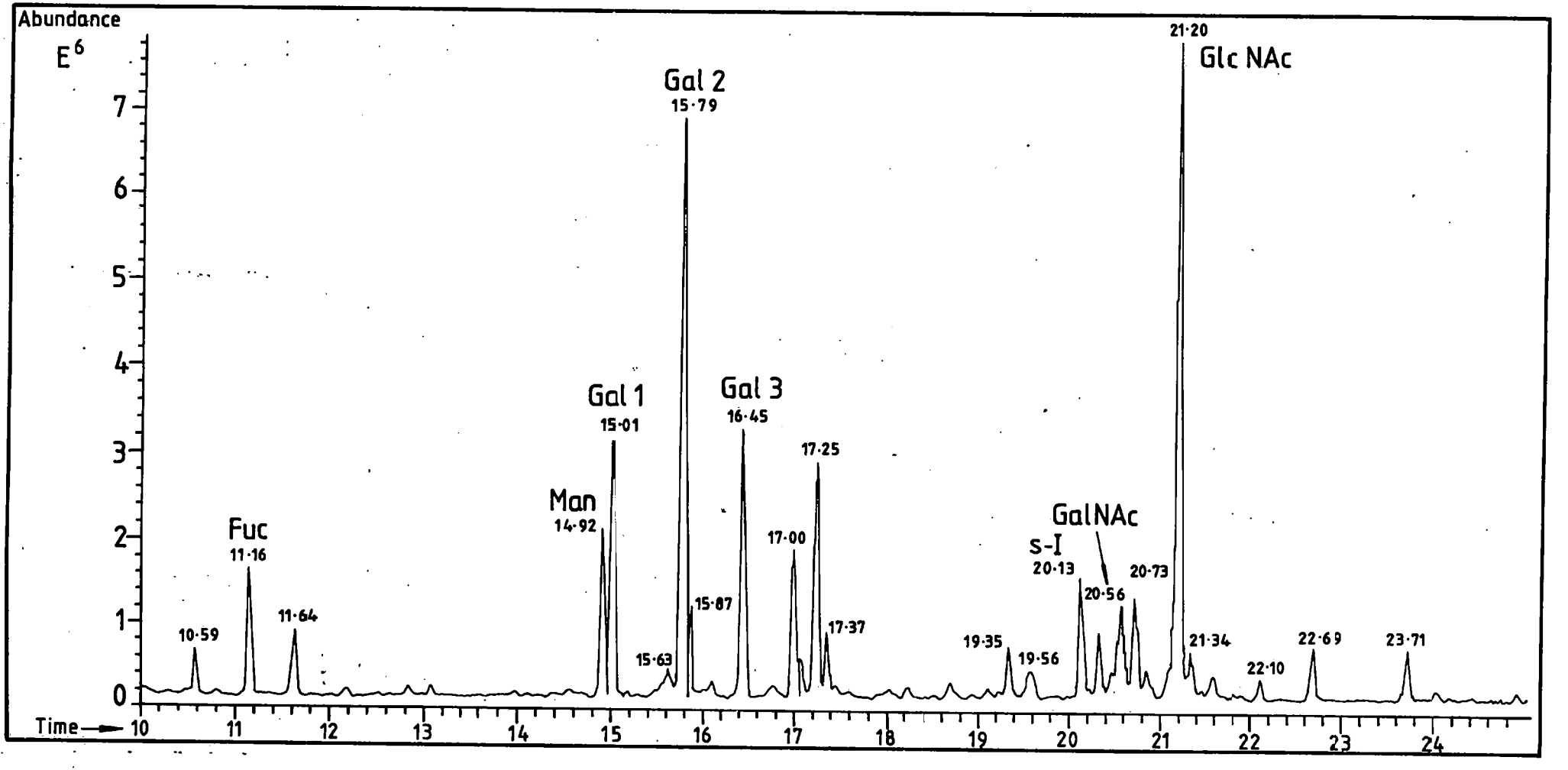


Figure 11c: GCMS composition analysis of glycans associated with CD59
Expansion of 0-24 minute section

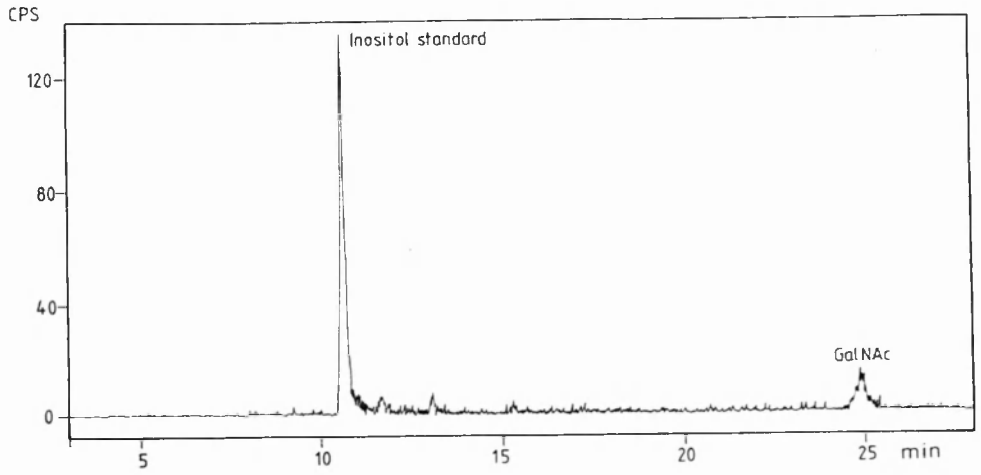


Figure 10b: Radio GC analysis of the reducing terminal glycan of the CD59 O-linked oligosaccharide.

Conclusion: CD59 contains the O-linked glycan: Gal β1- 3/4GalNAc

4.4 GCMS composition analysis of all CD59 sugars following methanolysis :

Figure 11a shows a typical cleavage pattern of a glycan anchor subjected to acid methanolysis. Figures 11b and c show the GCMS oligosaccharide composition analysis data following acid methanolysis of CD59. Figure 11b shows the full spectrum including the sialic acid peak, 11c is an expansion of the region containing fucose, mannose, galactose, N-acetyl galactosamine, and N-acetyl glucosamine. The peaks were assigned by analogy with known standards and the areas under each monosaccharide peak and the molar proportions of each of the species are shown in Table 1. The internal standard was *scyllo*-inositol which was present at 4nmoles.

Oligosaccharide	**Peak area**	**nmoles/50μg CD59**
N-acetyl galactosamine	7.79E7	6.1 nmoles
N-acetyl glucosamine	3.44E8	34.6 nmoles
mannose	6.64E7	5.2 nmoles
galactose	2.38E8	31.6 nmoles
fucose	5.14E7	8.2 nmoles
sialic acid	4.39E7	10.0 nmoles
scyllo-inositol	5.69E7	4.0 nmoles

Table 1 : *GCMS Composition analysis of the glycans associated with 50μg CD59*

The following conclusions can be drawn from the composition analysis of the CD59 glycans:

(i) The presence of GalNAc identified in the O-linked GalGalNAc glycan was confirmed.

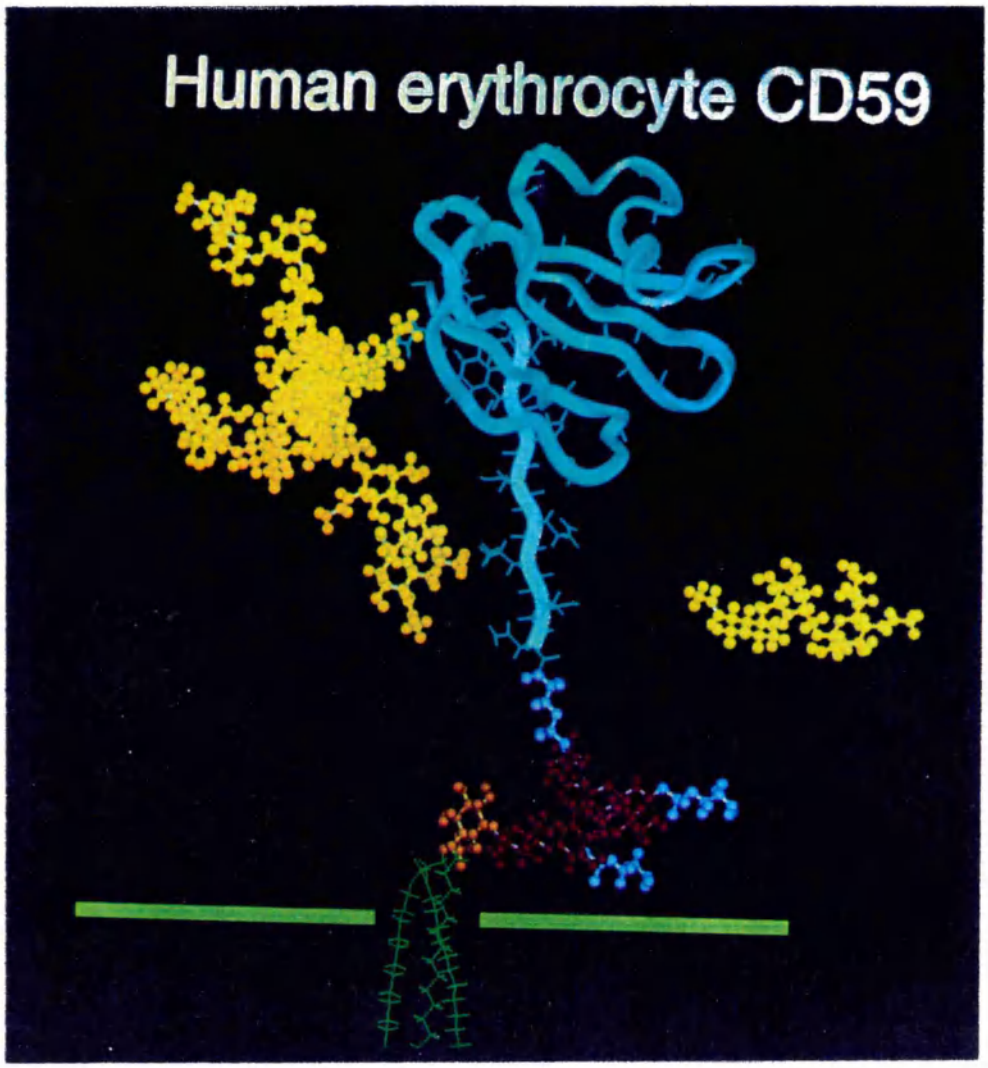


Figure 12: Molecular model of CD59 *showing the glycoprotein modelled with (i) a polylactosamine N-glycan (yellow) at Asn18, (ii) a glycan anchor with a tri-mannosyl core (red) (iii) an ethanolamine phosphate bridge at Man3 (mauve) and additional ethanolamine phosphate groups at Man1 and Man2 in the anchor (also mauve) (indicated by preliminary analysis - data not shown) (iv) with two lipids (dark green) attached to inositol (brown) C2 via phosphate and a third attached directly to the inositol ring at C1 (indicated by preliminary GCMS analysis and resistance to PIPLC) (v) the O-linked glycan (yellow) modelled alongside as the site of attachment to the protein is unknown. The protein structure (blue) is based on the 3D-NMR solution structure (Kieffer et al 1994).*

(ii) The high proportion of N-acetyl glucosamine and galactose compared with mannose is consistent with a high percentage of multi-antennary and polylactosamine type oligosaccharides.

(iii) The high percentage of fucose suggests outer arm fucosylation as well as core fucosylation.

(iv) The presence of sialic acid residues was confirmed.

4.5 Molecular modelling

4.5.1 The N-linked oligosaccharides may orient the N-terminal active site of CD59 towards the binding site on the C5b-9.

Figure 12 shows CD59 modelled with (i) a polylactosamine N-glycan at Asn18, (ii) a glycan anchor with a tri-mannosyl core (iii) an ethanolamine bridge at Man3 and additional ethanolamine groups at Man1 and Man2 in the anchor (indicated by preliminary analysis of the positions of the ethanolamine substituents - data not shown) (iv) with two lipids attached to inositol via phosphate and a third attached directly to the inositol ring (indicated by preliminary GCMS analysis and resistance to PIPLC). (v) the O-linked glycan modelled alongside as the site of attachment to the protein is unknown.

The N-linked oligosaccharides (size range 3-6nM) attached to CD59 (diameter approximately 12nM) at Asn18CysSer are close to the membrane surface. The glycans may orient the N-terminal active site of CD59 towards the binding site on the C5b-9 complex by restricting the conformational space available to the protein and limiting its interaction with the lipid bilayer.

4.5.2 *The asparagine-sugar amide linkage of CD59 is not protected from the proposed active sites of PNGase F by the local protein conformation.*

Figure 13 shows the crystal structures of CD59, RNase B and PNGase F. The structure of PNGase F is based on the crystal structure of Norris et al 1994. The protein is folded into two domains, each with an eight stranded antiparallel β-jelly roll configuration similar to the types of structures found in lectins. CD59 was modelled according to the crystal structure of Kieffer et al 1994. RNase B was modelled according to the crystal structure of Williams et al 1987. All three molecules are modelled to the same scale. Two active sites have been proposed for PNGase F: (i) Thr42, Lys44, Asp99 and Thr101 on strands D and G of domain 1 (ii) a site in a cleft at the opposite end of the molecule, flanked by five tryptophan residues and one phenylalanine residue. In contrast to the CD59 linkage the asparagine-sugar amide linkage in RNase was buried within the protein and not accessible to the active site of PNGase F. This is consistent with the experimental evidence in section 4.2 and in ch.4 which demonstrates that the N-linked glycans can be cleaved from native CD59, but not from undenatured ribonuclease B.

Figure 13: The crystal structures of CD59, RNase B and PNGase F

The structure of PNGase F is based on the crystal structure of Norris et al 1994. The protein is folded into two domains, each with an eight stranded antiparallel β-jelly roll configuration similar to the types of structures found in lectins. CD59 was modelled according to the NMR solution structure of Kieffer et al 1994. RNase B was modelled according to the crystal structure of Williams et al 1987. All three molecules are modelled to the same scale. Two active sites have been proposed for PNGase F: (i) Thr42, Lys44, Asp99 and Thr101 on strands D and G of domain 1 (ii) a site in a cleft at the opposite end of the molecule, flanked by five tryptophan residues and one phenylalanine residue. In contrast to the CD59 asparagine-sugar amide linkage, which is susceptible to the enzyme, the same linkage in RNase is protected by the protein and not accessible.

5. Discussion:

5.1 A comparison of CD59 and AChE suggests that structural features of the GPI anchors may be cell type specific (Fig. 14).

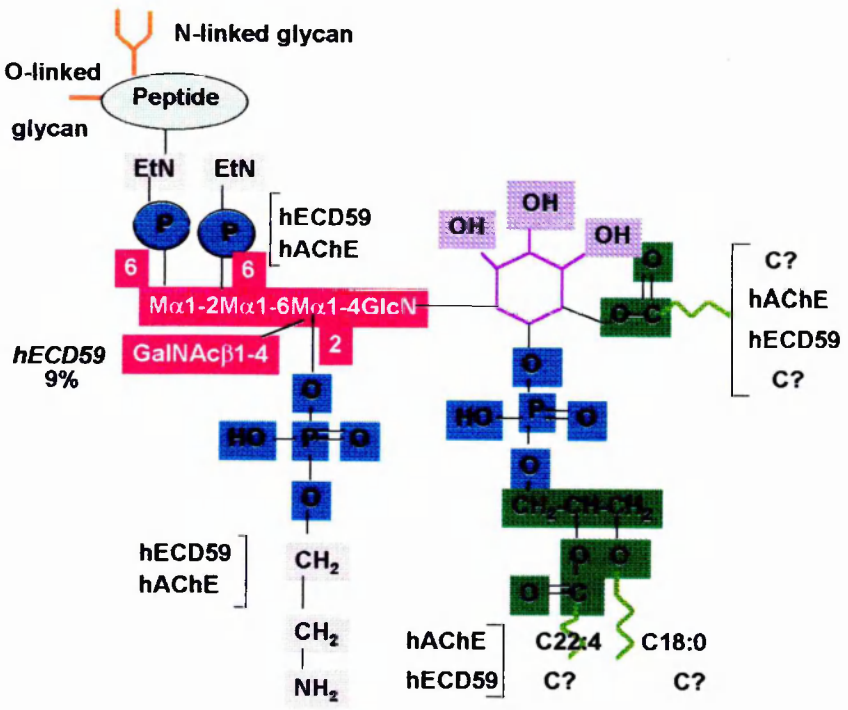
This study has enabled the first comparison of the glycosylation of two different GPI anchored proteins from the same cell. It raises the possibility that the glycosylation of GPI anchors may be cell type specific. 9% of erythrocyte CD59 is attached to the cell membrane by a GPI anchor in which the Manα1-2Manα1-6(GalNAcβ1-4)Manα1GlcNH$_2$ glycan links the protein to myo-inositol phosphate and the lipid. However 90% of the anchor glycans are of the Man α1-2Manα1-6 Manα1-4GlcNH$_2$ type. This is identical to the glycan in the GPI anchor which attaches to AChE to human erythrocytes (Deeg et al 1992).

In addition, unusually, AChE and CD59 both contain an additional lipid attached to the inositol ring, and preliminary data suggest that, in common with AChE, CD59 contains two additional ethanolamine phosphate groups attached to Man1 and Man2 in the glycan anchor. This may reflect the fact that, during biosynthesis, pre-assembled GPI anchors are attached to all proteins which contain the appropriate signal sequence. It may therefore be predicted that the GPI anchors attached to different proteins in the same cell will have common structural features. Subsequent processing of the tri-mannosyl glycan core is also controlled by the glycosylation machinery within the individual cell and may, in some cases, also be determined by the local conformation of the protein. More comparative studies are needed to explore the measure of control which the individual protein exerts over the glycosylation of its own anchor. The principle, however, has been well established for N-linked glycans where the further processing of the Glc$_3$Man$_9$ structure attached to all N-glycosylation sites is species, cell and site specific.

5.2 Glycosylation analysis may be a useful means of identifying the source of secreted glycoproteins.

In general, glycosylation analysis may be a useful means of identifying the source of urinary, serum or other secreted glycoproteins.

Glycans associated with CD59 in the urine may originate from both soluble and membrane bound CD59; in addition the sugars may be degraded during the processes of excretion suggesting that the structures shown in Tables 2a and 2b are minimum structures. Interestingly, the major portion of soluble urine CD59 (U-CD59) has been shown to contain a GPI moiety terminating with inositol and not with inositol phosphate (Nakano et al 1994). This explains the absence of the inositol phosphate attached lipid tails and gives insight into the origin of the secreted form of CD59. Comparison of the glycosylation of the erythrocyte CD59 anchor with that of urine CD59 (U-CD59) indicates that most U-CD59 is unlikely to be derived from erythrocytes. This is consistent with the fact that erythrocytes are processed mainly in the spleen and that, in contrast to erythrocyte CD59, U-CD59 contains no lipid directly attached to inositol.


N-linked glycan
O-linked
glycan
Peptide
EtN
EtN
P
P
6
6
hECD59
hAChE
Mα1-2Mα1-6Mα1-4GlcN
2
hECD59
9%
GalNAcβ1-4
O
HO
P
O
O
CH2
CH2
NH2
hECD59
hAChE
OH
OH
OH
O
O
C
C?
hAChE
hECD59
C?
O
HO
P
O
O
CH2-CH-CH2
O
O
O
C
hAChE
hECD59
C22:4
C?
C18:0
C?

Two studies have analysed the glycans attached to U-CD59. In the first (Table 2a) the heterogeneous set of glycans attached to Manα1-AHM of the dAR-uCD59 anchor included only 27.2% of the Man_3 structure found in erythrocytes:

Manα1-2Manα1-6: 28.1%
Manα1-2Manα1-2Manα1-6: 27.2%;
Manα1-2Manα1-6(GalNAcβ1-4): 33.9%;
Manα1-2Manα1-2Manα1-6(GalNAcβ1-4):10.8%

Table 2a: The glycans associated with the anchor from UCD59 (Nakano et al 1994)

In the second, soluble U-CD59 anchor glycans were also more heterogeneous those of the erythrocyte cell surface protein. However a different set of structures were identified which did not include the Man_3 glycan. The glycans attached to $ManGlcNH_2$*myo*-inositol assigned by mass spectrometry (and without full linkage information) are shown in Table 2b:

Man_2: 14%;
Man_2GlcNAc: 13%;
Man_3GlcNAc + Man_2GalGlcNAc:49%;
Man_3GalGlcNAc:15%;
Man_2(NeuNAcGalGlcNAc): 5%;
Man_3(NeuNAcGalGlcNAc): 4%

Table 2b: The glycans associated with the anchor from UCD59 (Sutton et al 1994).

The inconsistencies between these two sets of data may be due to different isolation procedures or to differences in the individuals from whom the samples were obtained, however neither set is fully compatible with the human erythrocyte CD59 analysis which contained 90% Man_3GlcNH. Thus, by an examination of the structures of the glycans, it is possible to conclude that the major part of CD59 in urine does not originate from erythrocytes. These data also imply that the glycosylation of GPI anchors is cell type specific.

5.3 Human erythrocyte CD59 contains an additional lipid chain substituted at inositol C2

Human erythrocyte CD59 is PIPLC resistant. The PIPLC cleavage reaction depends on the formation of a cyclic phosphate between the phosphate group at inositol C1 and the -OH group normally present at C2 (Fig.15); therefore it can be proposed that the third lipid identified in this study is attached to the inositol ring at C2. The presence of a third lipid chain provides an even more secure anchor into membrane making it unlikely that CD59 would be released either to allow it to re-enter the lipid bilayer at another point or to transfer to the membrane of another cell. This may be important if such a mechanism prevents CD59 solubilising in the membranes of non-host cells targeted for complement attack.

There is a continuing debate regarding the origin of soluble CD59. The study of U-CD59 suggests it is processed without the lipid tails and therefore normally secreted from cells. It has been suggested that soluble CD59 may also result from a cleavage of GPI anchored

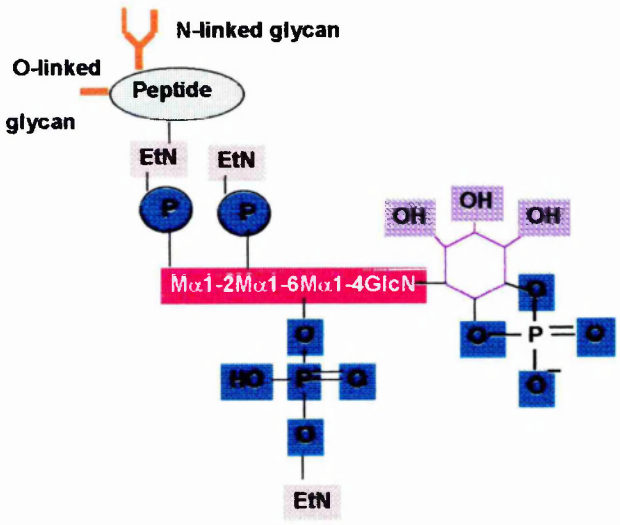


Figure 15: PIPLC cleavage of the lipids attached through ethanolamine phosphate at inositol C2 in GPI anchors. *This requires the formation of a cyclic phosphate involving positions C1 and C2. This is precluded by substitution of a third lipid at C1.*

molecules by endogenous phospholipases. The data from this study suggest that soluble CD59 is unlikely to result from a cleavage reaction which requires the normally anchored CD59 molecule to be temporarily out of the membrane. Interestingly, both Decay Accelerating Factor (DAF) (Walter et al 1990) and human AChE (Roberts et al 1989, Deeg et al 1992) also contain the third lipid substituted in the inositol ring.

5.4 The N-glycosidic linkage of erythrocyte CD59 is accessible to peptide N- glycanase F

Recent studies (Suziki et al 1993) have revealed a series of species and tissue specific PNGases active in mammalian cells. Some glycoproteins, for example plasminogen, contain aspartic acid in place of asparagine at some unoccupied glycosylation sites. This suggests that variable site occupancy may in some situations be the result of the removal of an attached glycan chain. Protection of the N-glycosidic linkage from cleavage or, alternatively, making it accessible to specific types of PNGase may therefore be important processing mechanisms. The enzyme is also widely used to release N-linked sugars from proteins for structure-function studies where it is important to recover an undenatured protein. In several cases, for example in RNase B (ch4), this has not proved possible. The enzyme may be used as a probe to compare the accessibility of the N-glycosidic linkage in CD59. In contrast to RNase B, the N-linked glycans were readily removed from the undenatured CD59 protein, suggesting that, in this case, the local conformation of the protein does not shield the N-glycosidic linkage from the active site of the enzyme. The reaction was explored by molecular modelling (fig.13) and the result is in contrast to ribonuclease B (ch.4) in which the asparagine-sugar amide linkage is protected from the proposed active sites of PNGase F by the local protein conformation.

5.5 68% of the glycans attached to CD59 are sialylated

Weak anion exchange chromatography and composition analysis using gas chromatography/mass spectrometry demonstrated that Hu erythrocyte CD59 is heavily sialylated. It contains 68% of 2,6- linked mono-, and a 2,3- linked di- and tri- sialylated glycans compared with 32% of neutral structures. This is consistent with the fact that most serum glycoproteins contain sialylated sugars. In contrast, human soluble CD59 expressed in CHO cells has no tri-sialylated glycans (ch.2A). UCD59 contains only 4% of monosialylated structures; 96% of the oligosaccharides are neutral. This may reflect the finding that it is asialo glycoproteins which are most rapidly removed from the circulation (for example by the asialo glycoprotein receptor in the liver) and excreted.

5.6 Human erythrocyte CD59 contains a high proportion of polylactosamine, multiantennary N-linked glycans

CD59 is glycosylated at Asn 18 with a range of smaller N-linked glycans including extended complex biantennary sugars (15.9gu, approximately 2,000D). However more than 40% are bi-, tri or tetra- antennary structures ranging from 20-30gu (approximately 2,500 - 4,000 D)

containing polylactosamine extensions. The molecular model (Fig. 12) suggests that the N-linked sugars may play a role by preventing interactions of hydrophobic regions on the protein with the lipid bilayer. This may be important for CD59 to move rapidly to the site of insertion of the C5b-8 MAC complex. The sugar may also orient the protein, increasing the affinity of CD59 for a site on C5b-8 and allowing it to compete effectively with C9. This is consistent with the finding that 88% of the ability of CD59 to restrict homologous lysis is lost when the glycan is removed. The composition analysis (Table1) confirms that CD59 contains a high proportion of N-acetyl glucosamine (34.6nmoles) and galactose (31.6nmoles) compared to mannose (5.2nmoles) indicative of polylactosamine extensions.

Once erythrocytes are released into the circulation they rapidly lose the ability to synthesise proteins, including glycosyl transferases and glycosidases. The glycosylation profile of CD59 is therefore a combination of structures: those synthesised by the original glycosylation machinery of normoblasts, and those which are put on during the changes which follow the extrusion of the cell nuclei.

5.7 The presence of an anchor may allow further processing of CD59 glycans

In contrast to erythrocyte CD59 soluble human CD59 expressed in CHO cells (without the signal sequence and therefore the anchor) contains 41% bi-antennary core fucosylated complex glycans (14.7gu), 24% of a biantennary structure with one lactosamine extension (18gu) and only 15% of larger structures. It is unclear why the CHO cell line normally restricts the size of the glycans although it has the machinery to add polylactosamines. In this case, two possible reasons for the extended structures are (a) the anchor decreases the rate of transport of the protein through the Golgi and, (b) the glycosylating enzymes are all membrane anchored to the Golgi membrane allowing the anchored erythrocyte CD59 to interact with the enzymes more efficiently.

5.8 Human erythrocyte CD 59 contains at least one O-linked glycan

No O-linked sugars have been reported previously on any CD59 molecule analysed to date. Asialo human erythrocyte CD59 failed to bind peanut agglutinin, a lectin which is specific for one of the core sequences for O-linked sugars Galβ1,3GalNAc (Ninomiya et al 1992). This result may suggest either that the O-linked GalGalNAc glycans detected on CD59 are not linked β1-3 or that the sugar is not accessible to the lectin while attached to the protein.

Although there is no peptide region with a high proportion of serine, threonine and proline residues which might constitute an O-linked domain, there are potential O-glycosylation sites at Thr15, 29, 51, 52 and 60, and at Ser21. Proline residues close to Ser or Thr are known to favour O-glycosylation, perhaps because the proline ring causes a local change in the conformation of the peptide allowing neighbouring amino acids side chains to become more exposed (ch.8 3.5). A possible site for O-glycosylation may therefore be Thr10 which is preceded by a proline at residue 9 and is close to proline 7. The possibility also remains that the O-link site may be in the C-terminal sequence absent in the recombinant protein.

In contrast to erythrocyte CD59 no O-links were detected in the glycans released from soluble human CD59 expressed in CHO cells following hydrazinolysis at 95^{o}C (ch2). Neither the presence nor absence of O-links was established for UCD59.

Interestingly, while the deduced amino acid sequence of CD59 predicts a protein core of approximately 10-11kD the molecular mass for de-N-glycosylated human erythrocyte CD59 is 14kD by SDS PAGE. Currently it is not known if the residual mass can be accounted for entirely by the GPI anchor or whether there is also some unidentified O-glycosylation.

It is also worth noting that recombinant HuCD59 expressed in CHO cells has an apparent molecular mass of 24kD. After removal of the N-linked sugar, the protein migrates with native human CD59 to 14kD (Ninomiya et al 1992). The molecular weight differences between the two glycosylated proteins may be attributed to cell specific glycosylation and, in contrast to the findings in this thesis, the data suggest that either both or neither of the CD59 molecules is O-glycosylated.

5.9 CD59 contains α1,6-linked sialic acids

Although the A2 and A3 populations of CD59 were susceptible, NDV did not reduce the A1 population to neutral. This suggests that the addition of a 1,6 sialic acid to one arm of the oligosaccharide prevents the addition of any more sialic acid residues to the developing glycan chain, resulting in an A1 subset of glycoforms. It also suggests that α2,6 sialic acid may not be added to a sugar which already has a sialic acid residue linked α2,3.

5.10 The presence of complex sugars at Asn 18 is essential for the function of CD59

De-N-glycosylation of human erythrocyte CD59 with N-glycanase resulted in an 88% reduction of the C5b-9 complement inhibitory activity of the protein (Ninomiya et al 1992). There was no change in activity following digestion of CD59 with neuraminidase, suggesting that the presence of sialic acid is not required for the protective function of the sugar. Moreover, CHO cells transfected with Hu CD59 grown in the presence of 1-DMM showed a comparable reduction in the C5b-9 inhibitory activity. This suggests that the presence of N-linked Man_9Glc_3 sugars at Asn18, which result from the inhibition of the normal processing pathways by DMM, an inhibitor of α-mannosidase, cannot substitute for the larger native complex type.

The role of the sugars remains to be elucidated; it has been established that they do not contribute to the affinity of CD59 for the C8 and C9 components of the C5b-9 complex. This is controlled by epitopes on C8 and C9 which become exposed following a conformational change which occurs within the MAC complex (Ninomiya et al 1992).

5.11 Conclusions:

Some of the structural details of the glycosylation of the GPI anchored erythrocyte glycoprotein, CD59, have been elucidated. 91% of the anchor glycans were of the Man_3 type, identical to those on erythrocyte AChE, consistent with the addition of a pre-assembled

anchor to both proteins in the ER. However, in contrast to AChE 9% of the core glycans attached to erythrocyte CD59 had been further processed by the addition of GalNAc. This suggests that the processing of the core glycans of GPI anchors may be cell type specific; this has been well established for N-linked glycans. In addition, an O-linked disaccharide was identified and >50% of the N-linked structures were of the sialylated, multi-antennary polylactosamine type. The glycans attached to erythrocte CD59 were different from the structures identified in two studies of urine CD59 indicating that UCD59 does not originate from erythrocytes. Molecular modelling suggests that the N-linked sugar, which is close to the membrane, may orient the active site of CD59 towards the C5b-8 receptor by restricting the conformational space available to the protein and preventing interactions with the lipid bi-layer. Interestingly, another membrane anchored protein, CD2, also contains an N-linked glycan near the membrane (ch.2). This may also serve to orient the CD2 molecule for its role in mediating cell/cell interaction, thereby suggesting that glycosylation may indirectly maintain the spacing between opposing cell membranes.

7. References to chapter 3:

Brooimans, R.A., van der Ark, A.A.J., Tomita, M., van Es, L.A., and Daha, M.R. (1992) Eur. J. Immunol. 22 791 CD59 expressed by human endothelial cells functions as a protective molecule against complement mediated lysis.

Cross, G. (1990) Ann. Rev. Cell Biol. 2-39 Glycolipid anchoring of plasma membrane proteins.

Deckert, M., Kubar, J., Zoccola, D., Bernard-Pomier, G., Angelisoval, P., Horejsi, V. and Bernard, A. (1992) Eur. J. Immunol. 32 2943-2947 CD59 molecule: a second ligand for CD2 in T-cell adhesion.

Deeg, M.A., Humphrey, D.R., Yang, S.H., Ferguson, T.R., Reinhold, V.N. and Rosenberry, T.L. (1992) J. Biol. Chem. 267 18573-18580 Glycan components in the glycoinositol phospholipid anchor of human erythrocyte acetyl cholinesterase.

Ferguson, M.A.J. (1991) Current Opinion in Structural Biology 1 522-529 Lipid anchors on membrane proteins.

Ferguson, M.A.J., Homans, S.W., Dwek, R.A., and Rademacher, T.W., (1988) Science, 239, 753-759 Glycosylphosphatidylinositol moiety that anchors *Trypanosoma brucei* variant surface glycoprotein to the membrane.

Ferguson, M.A.J., Homans, S.W., Dwek, R.A., and Rademacher, T.W., Anand, R., and Williams, A.F. (1988) Nature 333, 269-272 Complete structure of the glycosylphosphatidylinositol membrane anchor of rat brain Thy-1 glycoprotein.

Fletcher, C.M., Harrison, R.A., Lachman, P.J. and Neuhaus, D. (1994) Current Biology Structure 2, 185-199. Structure of a soluble, glycosylated form of the human complement regulatory protein CD59.

Hahn W.B., Menu, E. Bothwell, A.L.M., Sims P.J., Bierer, B.E. (1992) Science 256 1805-7 Overlapping but non-identical binding sites on CD2 for CD58 and for a second ligand CD59.

Holguin M.H., Fredrick, L.R., Bernshaw, N.J., Wilcox, L.A. and Parker, C.J. (1989) J. Clin. Invest. 84 7-17 Isolation and characterisation of a membrane protein from normal human erythrocytes that inhibits reactive lysis of the erythrocytes of paroxysmal nocturnal hemoglobinuria.

Hughes, T.R., Piddleston, S.J., Williams, J.D., Harrison, R.A. and Morgan, B.P. (1992) Biochem J. 284 169-176 Isolation and characterisation of a membrane protein from rat erythrocytes which inhibits lysis by the MAC of rat complement.

Kieffer, B., Driscoll, P.C., Campbell, I.D., Willis, A.C., van der Merwe, P.A. and Davis, S.J. (1994) Biochemistry 33, 4471-4482. Three dimensional solution structure of the extracellular region of the complement regulatory protein CD59, a new cell-surface protein domain related to snake venom neurotoxins.

Lachman, P.J. (1991) Immunology Today 12 312-315 The control of homologous lysis.

Low, M.G. (1989) Biochem. Biophys. Acta 988 427-454 The glycophosphatidyl inositol anchor of membrane proteins.

McConville, M.J. and Ferguson, M.A.J. (1993) Biochem. J. 294, 305-324. The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes.

Meri, S., Morgan, B.P., Davies, R.H., Daniels, M.G., Olavesen, H., Waldman, H. and Lachman, P.J. (1990) Immunology 71 1-9 Human protectin (CD59), an 18,000-20,000 MW complement lysis restricting factor, inhibits C5b-8 catalysed insertion of C9 into lipid bilayers.

Nakano, Y., Noda, K., Endo, T., Kobata, A. and Tomita, M. (1994) Arch. Biochem. Biophys. 311 117-126 Structural study on the glycosyl-phosphatidylinositol anchor and the asparagine-linked sugar chain of a soluble form of CD59 in human urine.

Norris, G.E., Stillman, T.J., Anderson, B.F. and Baker, E.N. (1994) Current Biology Structure 2:1049-1059 The three dimensional structure of PNGase F, a glycosyl-asparaginase from *Flavobacterium meningosepticum.*

Ninomiya, H., Stewart, B.H., Rollins, S.A., Zhao, J., Bothwell, A.L.M. and Sims, P.J. (1992) J. Biol. Chem. 267 8404-8410 Contribution of the N-linked Carbohydrate of Erythrocyte Antigen CD59 to its Complement Inhibitory Activity.

Roberts, W.L., Santikarn, S., Reinhold, V.N. and Rosenberry, T.L. (1989) J.Biol.Chem, 263 18776-18784 Structural characterisation of the glycoinositol phospholipid membrane anchor of human erythrocyte acetyl cholinesterase by fast atom bombardment mass spectrometry.

Sugita,Y., Nakano,Y., Tomita, M. (1988) J. Biochem. 104 633-637 Isolation from human erythrocytes of a new membrane protein which inhibits the formation of complement transmembrane channels.

Sutton, C.W., Cottrell, J.S., Meri, S., Lehto, T., Tynela, J., and Badman, M. (1994) Poster Communication Structural characterisation of the N-linked Glycosylation site and C-terminal GPI Anchor of soluble urine CD59 by MALD Mass Spectrometry.

Suziki, T., Seko, A., Kitajima, K., Inoue, Y. and Inoue, S. (1993) BBRC 194 1124-1130 Identification of peptide: N-glycanase activity in mammalian-derived cultured cells.

Taguchi, R., Fukahashi, Y., Ikezawa, H., Nakashima I. (1989) Biochem. Biophys. Res. Commun. 164 468-473 Analysis of PI (phosphatidyl inositol)-anchoring antigens in a patient of paroxysmal nocturnal hemoglobinuria (PNH) reveals deficiency of IF5 antigen (CD59), a new complement-regulatory factor.

Takizawa, H., Takashi, K., Murakami, T., Okada, N. and Okada, H. (1993) Eur. J. Immunol. 23 2714-2716 Rapid transformant selection by human complement using HRF20 (CD59) cDNA as a selection marker.

Tomita, M., Tobe,T., Choi-Miura, N., Nakano, Y., Kusano, M. and Oda, E. (1991) Abstracts XIVth Int. Complement Workshop p.233 Structural analysis of MACIF (CD59).

Townsend, R.R., Hardy, M.R., and Lee, Y.C. (1989). In Methods in Enzymology Vol. 179 (ed. V. Ginsburg), p 779.

van der Merwe, A., Barclay, N., Mason, D., Davies, E., Morgan, B.P., Tone, M., Krishnam, A.K.C., Lanelli, C., and Davis, S. (1994) Biochemistry 33 10149-10160. Human cell adhesion molecule CD2 binds CD58 (LFA3) with a low affinity and extremely fast dissociation rate, but does not bind CD48 or CD59.

Walsh, L.A., Tone, M., and Waldman, H. (1991) Eur. J. Immunol. 21 847-850 Transfection of human CD59 complementary DNA into rat cells confers resistance to human complement.

Walsh, L.A., Tone, M., Thiru, S., Waldman, H. (1992) Med. Per. 55a 213-220 The CD59 antigen - a multifunctional molecule.

Walter, E.I., Roberts, W.L., Rosenberry, T.L., Ratnoff, W.D. and Medof, M.E. (1990) J. Immunol. 144 1030-1036 Structural basis for variations in the sensitivity of human decay accelerating factor to phosphatidyl inositol-specific phospholipase C cleavage.

Williams, R.L. & Greene, S.M., & McPherson, A. (1987) J. Biol. Chem. 263 16020-16031 The crystal structure of ribonuclease B at 2.5A resolution.

Zhao, J., Rollins, S.A., Maher, S.E., Bothwell, A.L.M., Sims, P.J. (1991) J. Biol.Chem. 266 13418-13422 Amplified gene expression in CD59 transfected chinese hamster ovary cells confers protection against the membrane attack complex of human complement.

Chapter 4
The Structure and Function of the Glycoform Populations of Ribonuclease B

Publications associated with this chapter:

1. Rudd, P.M., Scragg, I.G., Coghill, E. & Dwek, R.A. (1992) Glycoconjugate Journal 9, 86-91 The Separation and Analysis of the Glycoform Populations of Ribonuclease B using Capillary Electrophoresis.
2. Rudd, P.M., Joao, H.C., Coghill, E., Fiten, P., Saunders, M.R. Opdenakker, G. and Dwek, R.A. (1994) Biochemistry 33, 17-22 Glycoforms modify the dynamic stability and functional activity of an enzyme.
3. Rudd, P.M., Woods, R.J., Wormald, M.W., Opdenakker, G., Downing, A.K., Campbell, I. D. and Dwek , R.A. (1995) Biochim. Biophys. Acta - in press - The effects of variable glycosylation on the functional activities of ribonuclease, plasminogen and tissue type plasminogen activator.

Acknowledgements:

1. The initial part of this study was carried out in 1991 immediately prior to registration for this thesis and has been published (Rudd et al 1992). It is summarised briefly only as the background to the work which followed.
2. Comparisons of the dynamic stability and functional activity of the RNase B glycoforms were carried out in collaboration with Dr. Heidi Joao (Glycobiology Institute), Prof. Ghislain Opdenakker and Dr. Pierre Fiten (University of Leuven); the work is briefly summarised to set the study of RNase in this thesis into context.
3. Molecular modelling was carried out in collaboration with Dr. Martin Saunders (SmithKline Beecham).
4. Assistance by Ms. Eva Coghill in the analysis of the enzyme digestions by capillary electrophoresis is acknowledged.

Abbreviations: CE: capillary electrophoresis; RNase: ribonuclease; DC: diffusion coefficient; RNase B Man 5: the glycoform of RNase containing 5 mannose residues; tPA: tissue plasminogen activator; HPAEC: High Performance Anion Exchange Chromatography.

Abstract

Glycoproteins generally consist of a mixture of glycoforms, however the reasons for such diversity are not well understood. This study explores the possibility that the protein structure may control the accessibility of the N-linked oligomannose sugars to processing enzymes and, conversely, that the presence of different oligosaccharides may modulate the recognition, stability and functional activity of some of the glycoforms of RNase B. Intact ribonuclease B was resolved by borate capillary electrophoresis into five populations (RNase Man 5-9) according to the particular oligomannose structure associated with each glycoform. Alterations in the composition of the glycoform populations during digestion of ribonuclease B with A. Saitoi α1,2 mannosidase were then monitored. The terminal α1,2 linked mannose residue of the Man 6 glycoform was more resistant to the enzyme than the α1,2 linked mannose residues of the higher glycoforms, suggesting that the extent to which the protein can shield the sugar chain is variable and depends on the proximity of the protein to each oligosaccharide residue. Digestion of the free oligosaccharides under the same conditions, monitored by High Performance Anion Exchange Chromatography (HPAEC), revealed that the rate of cleavage of the free sugars was 100 times faster, suggesting a role for the protein in shielding the oligosaccharides from processing enzymes during biosynthesis. To determine whether the sugars confer stability on the protein, the Man 5,1 and 0 glycoforms were tested for resistance to pronase. All the glycans gave equal protection to RNase B compared with the unglycosylated protein. A polyclonal antibody raised against RNase A showed a reduced affinity both for the Man5 glycoform and for the native RNase B but not for RNaseMan1 demonstrating that some N-linked sugars can shield sites on the protein from antibody recognition. Pure individual glycoforms tested in a functional assay using double-stranded RNA substrate showed nearly a 4-fold variation in activity; molecular modelling suggested that steric factors may play a role in modulating this interaction and that the sugars may variably disrupt one or more of the salt bridges between anionic residues in RNaseB and phosphate groups in the double stranded RNA substrate.

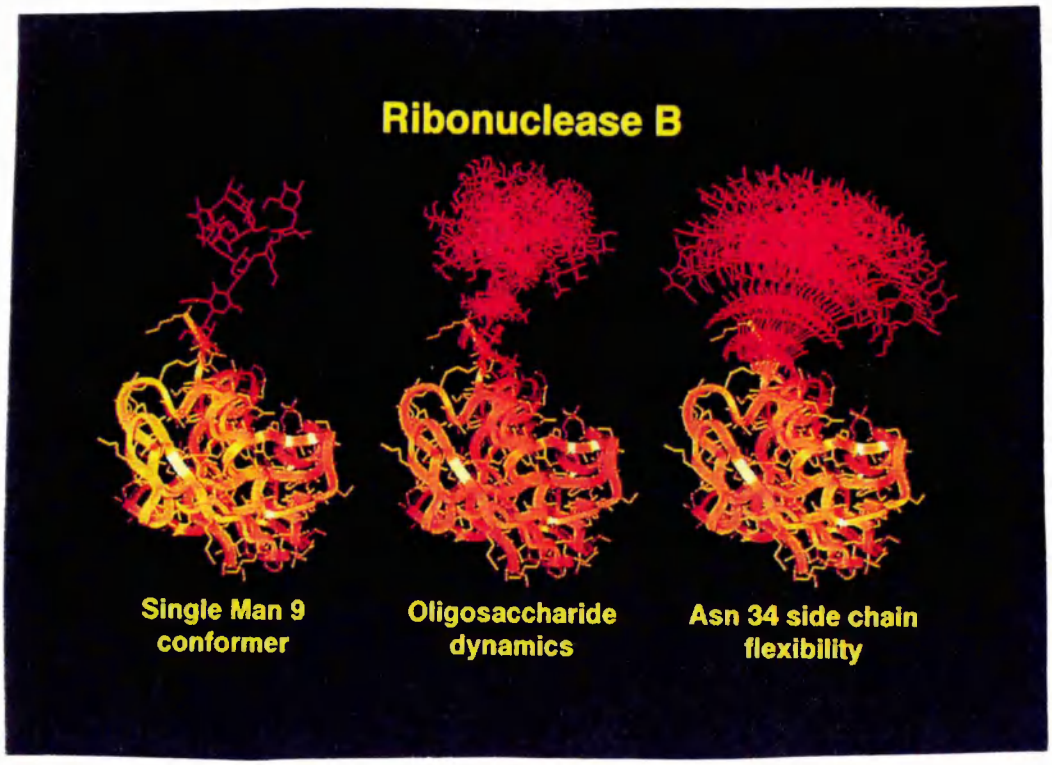


Figure1a: A. A single Man9 conformer of RNase based on the 2.5A X-ray crystal structure. B. A single Man9 conformer of RNase showing the dynamics of the oligosaccharide based on an overlay of 10 oligosaccharide conformations. C. The effect of Asn34 side chain flexibility on the orientation of the oligosaccharide in the Man9 glycoform of RNase B. (Taken from Woods et al 1994 and Rudd et al 1995). The protein structure is based on the crystal structure (Williams et al 1987).

Amino acid sequence of RNase

KETAA AKFER QHMDS STSAA SSSNY CNQMM	1-30
KSRNL TKDRC KPVNT FVHES LADVQ AVCSQ	31-60
KNVAC KNGQT NCYQS YSTMS ITDCR ETGSS	61-90
KYPNC AYKTT QANKH IIVAC EGNPY VPVHF	91-120
DASV	121-124

N-linked glycosylation site: Asn34LysThr

Active site residues include His12, Lys41, His119

Disulphide bonds: 26-84, 40-95, 58-110, 65-72

Molecular weight of unmodified chain: 13,690

Figure 1b

The Structure and Function of the Glycoform Populations of Ribonuclease B

1. Background

1.1 Proteins may modulate the biosynthesis of N-linked glycans

In chapter 2 it was suggested that the protein plays a role in modulating the biosynthesis of N-linked glycans. For example, site 65 on CD2 is glycosylated exclusively with oligomannose structures, and examination of a molecular model of CD2 indicated that Asn65 may lie in a cleft in the protein. This suggests that, in some cases, the local protein conformation may restrict the structure of its own glycoforms to the oligomannose class by shielding developing glycans from mannosidase and N-acetylglucosaminyl transferase processing enzymes. To explore such a possibility and, conversely, to probe the extent to which the particular oligosaccharide structures associated with single glycoform populations can specifically modify accessibility, stability and functional activity, a study was made of bovine pancreatic ribonuclease.

1.2 Structure of bovine pancreatic ribonuclease

This single domain, globular protein (Fig.1a,b) is a mixture of unglycosylated RNase A (13,690kd) and a collection of glycoforms (designated RNase B Man-5,-6,-7,-8 and -9) in which the oligomannose series, Man-5 to -9 (Fig.2), is associated with the single N-glycosylation site at Asn34LeuSer. Although there have been many studies comparing the activity of glycoproteins with their deglycosylated counterparts, in general the significance of oligosaccharide diversity at one site is not known. For example, in this thesis functional studies of tissue plasminogen activator (t-PA) (ch.7) have shown that the rate of fibrin dependent plasminogen activation spans a range which depends on the glycosylation site occupancy of both t-PA and its substrate, plasminogen. At the extremes the fibrin dependent activity for Type II t-PA (two N-linked glycans) with Type 2 plasminogen (one O-linked glycan) is 2-3 times the value for Type I t-PA (three N-linked glycans) and Type 1 plasminogen (one N- and one O-linked glycan) (Wittwer et al 1989; Mori et al 1994). However, it is not known whether, within this framework, glycoform diversity at each individual site can provide a spectrum of activity, allowing the cell a means of achieving a finely regulated response to its environment. Such studies have not been possible previously because the technology required to separate individual glycoforms of a glycoprotein has not been available. However, the development of capillary electrophoresis to resolve the five glycoforms of bovine pancreatic RNase B (Rudd et al 1992) has allowed the preparation and analysis of individual pure glycoforms of a glycoprotein making it possible to address the fundamental question of the consequences of glycosylation changes at a single site.

Many physical parameters have been determined for bovine pancreatic RNase; in particular the crystal structures of both RNase A and B are available (Wlodawer et al 1988; Williams et al 1987). The active site, centred around His12, His 119 and Lys41, has been well defined (Shall and Barnard 1969), as have the anionic binding sites involved in its interaction with RNA

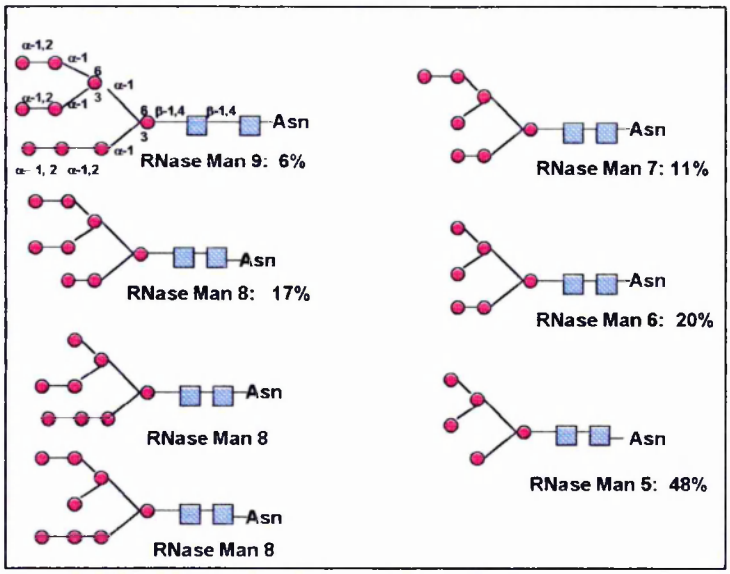


Figure 2a: *The oligomannose glycoforms of RNaseB. Most of these glycoforms exist as a number of isoforms which differ in branching and linkage. To illustrate this three alternative structures of the Man8 glycoform are shown. The relative proportions of the five populations were determined by capillary electrophoresis.*

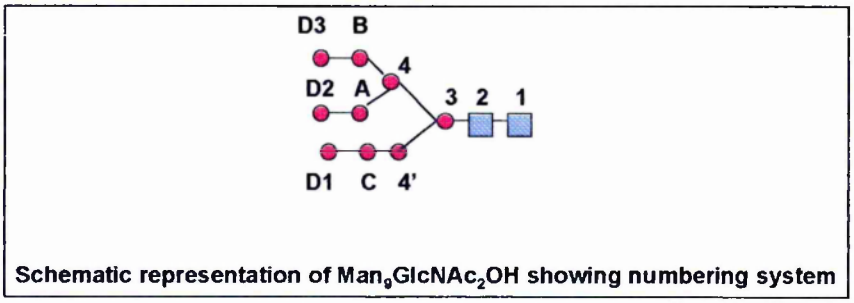


Figure 2b

(McPherson et al 1986). Importantly, X-ray crystallography (Williams et al 1987), circular dichroism and optical rotatory dispersion (Puett 1973) and 1D- and 2D-1H-NMR (Joao et al 1992) have demonstrated that the secondary and tertiary structure of RNase is unaffected by glycosylation and therefore the overall 3-D conformation of individual RNase glycoforms is the same.

1.3 The function of ribonuclease

Mammalian ribonucleases regulate many important biological functions by catalysing the degradation of single- and double- stranded ribonucleic acids (RNAs). RNase secreted from the pancreas catalytically degrades RNA in the gut, and is associated with pancreatic cancer (Reddi and Holland 1976). RNase is expressed in other normal and tumour cells (e.g. osteosarcoma, leiomyosarcoma; P. Fiten et al unpublished data). RNases are cytotoxic (D'Allessio et al 1991; Wool 1984), moreover RNases interfere with the double-stranded replicative intermediates of many viruses, with duplex-based latency of Herpes viruses (Stevens et al 1987), with viral capsid formation (Fisher and Johnson 1993) and with double-stranded RNA-mediated induction of cytokines such as interferon (Van Damme et al 1985), IL-6 (Van Damme et al 1987) and chemokines (Van Damme et al 1991). Ribonuclease-nucleic acid interactions have been extensively studied, but the contribution of the oligosaccharides has not been previously addressed.

2. Introduction

This study explores (a) the extent to which the protein structure protects the oligomannose sugars associated with the different glycoforms of RNase B from modifications by A. Saitoi mannosidase enzyme (b) whether the oligomannose sugars variably shield protease susceptible sites on the protein surface (c) whether the oligomannose sugars on RNaseB variably shield antigenic sites on the protein from a polyclonal antibody raised against RNase A (d) the possibility that the oligomannose sugars may variably modulate the activity of RNase towards double stranded RNA by steric hindrance.

3. Materials and Methods

3.1 Capillary Electrophoresis of RNase A and B

A 72cm fused silica capillary, ID 75µm, was used; voltage conditions were 1kv for 1min, 20kV for 9min, 30^{0}C, wavelength for detection 200nm, injection time 1.5s, in 20mM sodium phosphate, 50 mM SDS, 5mM sodium tetraborate pH 7.2, on a Beckman P/ACE system with GoldTM Software. HPLC used a Waters System with a TSK 3000 gel filtration column equilibrated in potassium phosphate buffer (0.1 M pH 7.2).

3.2 Purification of RNase A and B

Commercially available RNase A is frequently contaminated with RNaseB and vice-versa. Concanavalin A lectin binds proteins containing oligomannose structures and can therefore be used to selectively bind the glycosylated variants of RNase. Where necessary, batches of RNase purchased from Sigma Chemical Company were typically further purified as follows:

Purification of RNase A

Concanavalin A sepharose (Pharmacia) was pre-equilibrated with (i) 0.1M HCl followed by (ii) NaOAc (5mM;pH5.2) containing $CaCl_2$ (1mM) and $MgCl_2$ (1mM) and 5mM α-methyl mannoside (iii) NaOAc (5mM;pH5.2) containing $CaCl_2$ (1mM) and $MgCl_2$ (1mM) (Buffer A). 1mg of RNase A (Sigma) was applied to the column in the equilibration buffer. The unbound fraction (RNase A) was collected, dialysed into water and checked for purity by CE and SDS PAGE. 0.47mg was recovered (47% yield). No detectable amounts of Con A were leached from the affinity column.

Purification of RNase B

Concanavlin A was pre-equilibrated as above and RNase B (1mg Sigma) was applied to the column as described for RNase A. The column was washed with buffer A containing 0.1M NaCl to remove non-specifically bound molecules. RNase B (0.64mg) was eluted with 200mM α-methyl mannoside in the same buffer, dialysed into water and checked for purity by CE and SDS PAGE. Con A, which occasionally leached from the column in significant amounts, was removed by HPLC gel filtration using a TSK 3000 analytical HPLC gel permeation chromatography (GPC) column equilibrated in 0.1M potassium phosphate pH7.2.

3. 3 Enzymatic modifications of RNase B glycoforms:

RNase A and B (Sigma Chemical Co.) were purified to electrophoretic homogeneity by Concanavalin A sepharose affinity chromatography and HPLC GPC.

Individual glycoforms were prepared by enzymatic modification of RNase B. All the digestions were monitored by capillary electrophoresis; in the case of RNase Man-1 the reaction was carried out in the CE carousel and samples were injected directly into the capillary from the reaction mixtures, eliminating the need to take aliquots and stop the reactions. All peaks were assigned by co-injection of the digestion mixtures with the standard RNase B ladder preparation. This was obtained by mixing digestion products taken at different time points during the Jack Bean α-mannosidase digestion to obtain a mixture of glycoforms from Man9-Man1. The glycoforms were purified from exoglycosidase enzymes by gel permeation HPLC before analysis. The final samples were electrophoretically pure by CE.

Enzyme digests (see fig. 4b):

(i) RNase B Man-5: RNase B was digested to a single population of RNase B Man-5 with *A. Saitoi* α 1,2 mannosidase at an enzyme:substrate ratio of 25mU :1mg in 1ml 50mM sodium acetate pH 5 in 96h at 37^{o}C.

(ii) *RNase B Man-1*: RNase B was converted to a single population of RNase B Man-1 with Jack Bean α-mannosidase (5U:1mg in 1ml 10mM sodium citrate pH 4.5 / 0.2mM Zn acetate) in

95 min at 30°C. A slower digestion was performed to reveal details of the early part of the reaction used 2U:1mg in 1ml 10mM sodium citrate pH 4.5 / 0.2mM Zn acetate).

(iii) *RNase B Man-0*: RNase B Man-1 was converted to a single population of RNase B Man-0 with the β1-4 mannosidase *Helix pomatia* in 20h at 37°C in 100mM sodium acetate pH4, with an enzyme:substrate ratio 10U : 50mg /ml.

(iv) *RNase B GlcNAc:* RNase B was converted to a single population of RNase BGlcNAc with EndoH at 1U/mg RNase B in 100mM sodium citrate buffer pH 5.5 for 16h at 37°C.

(v) Digestion with PNGase F (20U/mg) in 20 mM sodium phosphate pH 7.5 containing 50mM EDTA for 18h at 37°C.

3.4 Pronase digestion

In three separate experiments RNase A together with RNase Man-5, Man-1, or Man-0 were incubated with pronase in the CE carousel at 30°C +/- 3°C. The enzyme to substrate ratio was 7.7U:10mg and the buffer was 50mMTris/HCl pH 7.5. Digestions were monitored continually by direct injection into the capillary and each incubation was repeated at least three times. The data were obtained by integration of the appropriate peaks. Similar digestions of the glycoforms in pairs were carried out under the same conditions.

3.5 Preparation of polyclonal α-RNase A

Purified RNase A was injected into a rabbit (20μg in complete Freund's adjuvant on day 1, 20 μg in complete Freund's adjuvant after 2,4,6 and 8 weeks, and pure antigen after 10 weeks). Blood (10ml) was collected after 11 weeks. The serum (2ml) was removed from the clot and freeze dried (Rega Institute, Leuven). The serum was reconstituted in 2ml of PBS and applied to a protein A/Sephadex G25 column equilibrated in 0.14M potassium phosphate pH8. IgG was eluted with glycine pH 2.8 and checked for purity by HPLC. 8mg of IgG were recovered.

3mg of purified RNase A were bound to a sepharose CL4B affinity gel (0.25ml). The purified IgG (8mg) in PBS was recycled for 1 hour at room temperature through the affinity matrix in tandem with 1ml of Sephadex G25 (Pharmacia) to lower the pH immediately after elution. After extensive washing with PBS, bound material was eluted with glycine pH 2.8 and the pH immediately adjusted to 7. 0.4 mg of IgG was recovered in the bound fraction.

3.6 Preparation of RNase A affinity column

700μg of freeze dried CNBr activated Sepharose 4B were swollen for 15 minutes in 1mM HCl and washed with 2 ml of the same solution (0.25ml gel). RNase A (3mg) was dissolved in the coupling buffer ($NaHCO_3$ [0.1M, pH8.3] containing NaCl [0.5M]. The protein solution was mixed with the gel suspension for 4 hours at room temperature after which time approximately 95% of the protein had bound to the gel. Remaining active groups were blocked with 0.2M glycine pH 8 and the excess protein washed away with coupling buffer followed by acetate buffer [0.1M, pH4] containing NaCl [0.5M] followed by a further wash with coupling buffer. The blocking agent was removed by the coupling buffer. The ratio of RNase A to gel was 11.4mg/ml.

3.7 ELISA assay

Microtitre plates were coated with the following mixtures of RNase A and RNase B glycoforms:

(1) RNase A 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% and
RNase B (Man5-9) 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 0%

(2) RNase A 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% and
RNase Man5 100, 90, 80, 70. 60. 50, 40, 30, 20, 10, 0%

(3) RNase A 0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100% and
RNase Man1 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 0%

All solutions were 5μg/ml in PBS

Appropriate volumes of each solution were mixed in the microtitre plate wells (total volume 200μl) and incubated overnight at room temperature to coat the wells with different mixtures of the RNase A and RNase B glycoforms as shown above. The plates were washed with 3x1% ovalbumin / 0.05% Tween in PBS and blocked with 1%BSA/PBS for 2 hours at room temperature. After washing with 3x1% ovalbumin / 0.05% Tween in PBS, rabbit IgGα-RNaseA (10ng/ml) was incubated in each well for 3 hours at 37°C (having previously established that this was a saturating level of antibody). The bound IgGα-RNaseA was detected with peroxidase labelled (goat) α-rabbit IgG by incubating 200μl of a 1:1000 dilution in each well for 2 hours at 37°C. The antibody was visualised by incubating with orthophenylene diamine. The reaction was stopped with H_2SO_4 and the result read at 405nm. The results are the average of eight runs / sample.

4. Results

4.1 The protein structure may protect the RNase B glycans from A.Saitoi mannosidase

The oligomannose sugars linked to RNase B were released by hydrazine, reduced and radiolabelled with tritium and digested to a single Man5 structure with *A.Saitoi* α1,2 mannosidase. The rate of digestion was monitored by High Performance Anion Exchange Chromatography (HPAEC) using a radioactive detector. The natural population of RNase B glycoforms (fig. 4a) was digested under identical conditions to a single population of the Man5 glycoform (Fig. 4b). In this case the digestion was monitored by CE equipped with a UV detector reading at 200nM. The time taken to digest 33 nmoles of the released oligomannose glycans at an enzyme:substrate ratio of 1:12.5mU was 34 minutes (Fig. 3a); the reaction was over 100 times faster than the digestion of the same sugars attached to 33 nmoles of RNase B which took more than 72 hours (Fig. 3b). The data in Fig. 3b also show that the Man 6 glycoform is more resistant to the enzyme than the higher glycoforms, Man7-9. Reaction rates depend on several factors, some of which, such as temperature (37°C), pH (5.0), concentration

(1mmolar) buffer (50mM sodium acetate) and enzyme:substrate ratio (1: 12.5mU) were held constant. Two major variables, which could not be controlled since they were characteristic of the two species being compared, were (i) the different diffusion rates of the sugars and the glycoprotein and (ii) the steric effects of the protein on the accessibility of the sugars in intact RNase B.

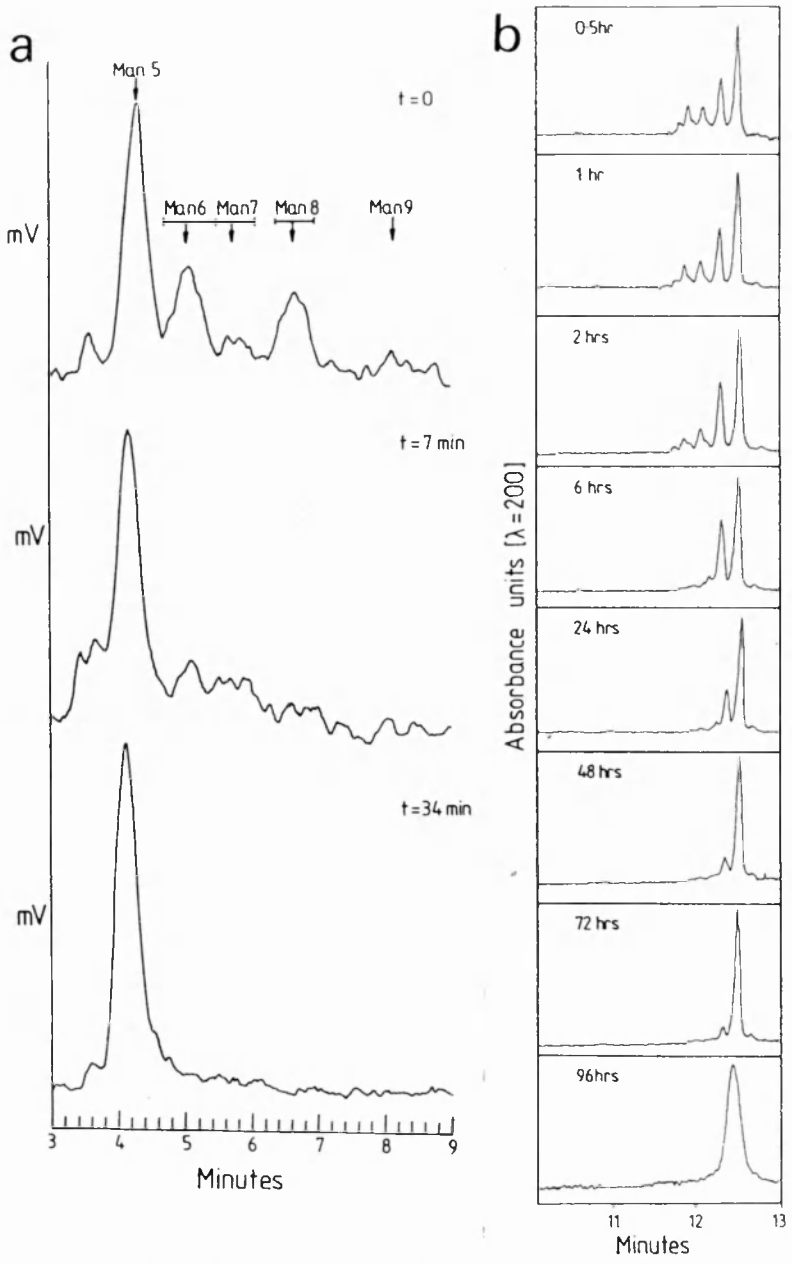


Figure 3: The rate of A.Saitoi mannosidase digestion of the (a) oligomannose glycans released from RNase B compared with (b) the rate of digestion of the same sugars attached to RNase B *(Rudd et al 1992)*

(a) Anion exchange chromatogram (Dionex CarboPac PA1) of (i) the reduced and radiolabelled N-linked oligosaccharides released by hydrazinolysis from ribonuclease B (ii) the same oligosaccharide library after digestion for 7 minutes with A. Saitoi α1,2 mannosidase (iii) after 34 minutes all oligomannose structures have been reduced to Man5, detected by the ^{3}H label at the reducing terminus.

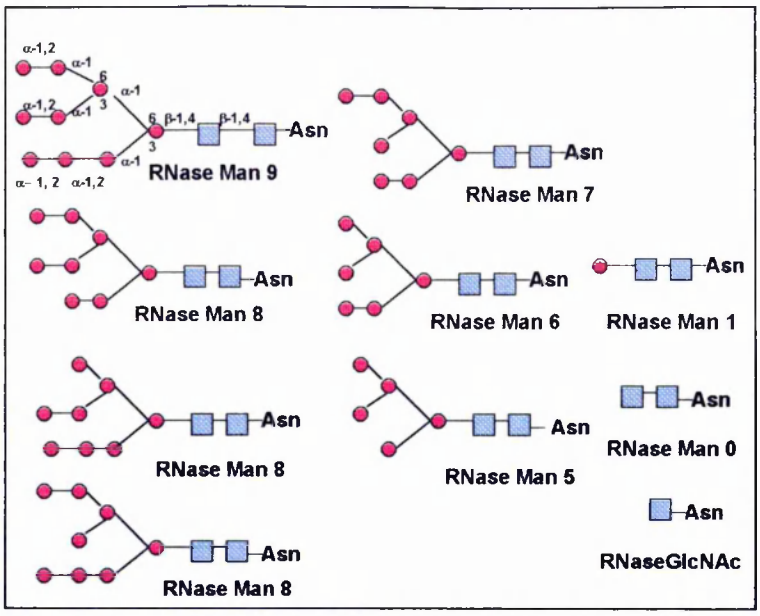


Figure 4a: The natural glycoforms of RNase B (RNaseMan9-5) and the reaction products (RNaseMan 5, 1,0 and GlcNAc) from the digestion of the oligomannose glycans on intact bovine pancreatic RNaseB. Most of these glycoforms exist as a number of isoforms. To illustrate this three alternative structures of the Man8 glycoform are shown.

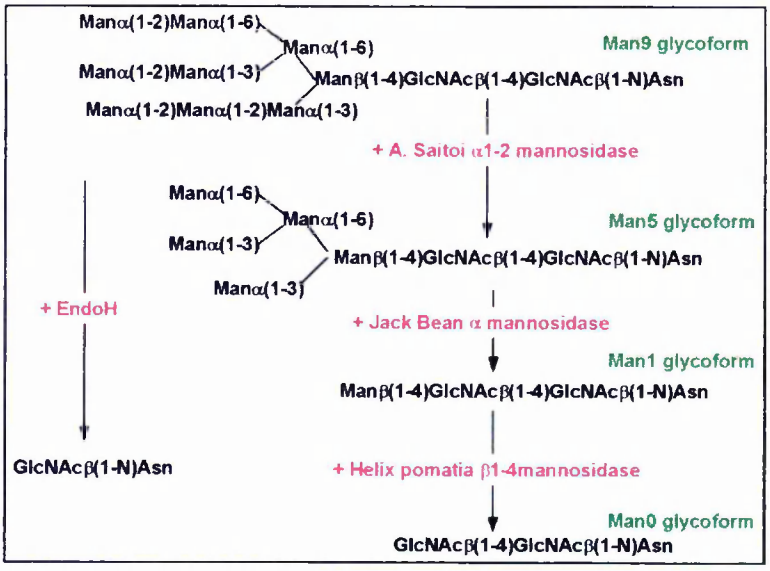


Figure 4b: Four enzymes were used to obtain the individual glycoforms (a) A. Saitoi α1,2 mannosidase yielding RNaseMan5 from the natural mixture of glycoforms (b) Jack Bean α1,2,6>3 mannosidase yielding RNaseMan1 from the natural mixture of glycoforms or from Man5 (c) Helix pomatia β1,4 mannosidase yielding RNaseMan0 from RNaseMan1 (d) EndoH yielding RNaseGlcNAc from the natural mixture of glycoforms.

(b) Capillary electrophoresis profiles of bovine pancreatic RNase B showing the time course for the digestion of the glycoprotein with A.Saitoi α1,2-mannosidase. Absorbance at 200nm is plotted against elution time for time points between 0 and 25h. This demonstrates that the five glycoforms (Man5:48%, Man6:20%, Man7:11%, Man8:17%, Man9:6%) initially present are reduced by the enzyme to a single structure, the one having the mannose 5 sugar. The Man6 glycoform is more resistant to the enzyme than the other higher glycoforms.

4.2 Enzymatic modifications of RNase B to obtain electrophoretically pure glycoforms.

Recent advances in capillary electrophoresis have enabled each of the oligomannose glycoforms of RNase B to be resolved at the protein level in their correct molar proportions. This has made it possible to analyse glycoform population changes directly and therefore to prepare a number of single, electrophoretically pure glycoforms for testing in the systems described in sections *5.3 - 5.6*. RNase Man-5, Man-1 and Man-0 were prepared by modification of the naturally population of RNase B glycoforms (RNase Man5-9) with the exoglycosidase enzymes *A. Saitoi* α-(1,2) mannosidase, Jack Bean α-mannosidase, *Helix pomatia* β1-4 mannosidase. RNaseGlcNAc was prepared using recombinant endoglycosidase H from *Streptomyces plicatus.* The digestion scheme is described in figure 4b and the results are shown in figure 5.

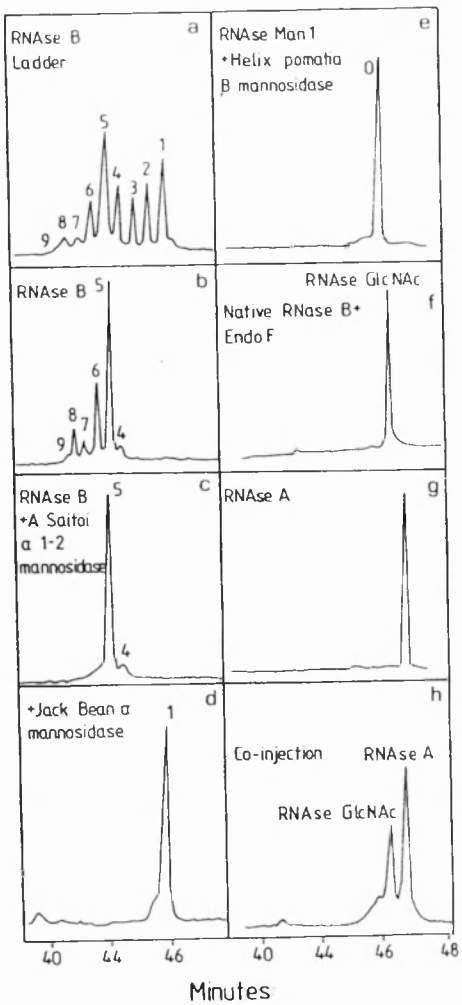


Figure 5: Preparation and CE analysis of individual glycoforms of RNase.

a. RNase B glycoform ladder (Man1-Man9) used to assign structures to the peaks obtained by CE analysis of the enzyme digests. To assign structures the standard mixture was co-injected with the samples.

b. Natural population of RNase Man-9,-8,-7,-6 and -5 glycoforms from bovine pancreatic RNase B which was the starting material for the biosynthesis of the individual Man-5, Man-1 and Man-0 glycoforms.

c. RNase-Man-5, prepared by reducing the whole natural glycoform population of RNase B to a single population with A. Saitoi α-(1,2) mannosidase which cleaves Man α(1-2) residues.

d. RNase Man-1, the product of the digestion of RNase B with Jack Bean α-mannosidase. This enzyme releases Man α(1-2), Man α(1-3) and Manα(1-6) residues from the natural mixture of RNase B glycoforms to yield the single Man-1 glycoform, .

e. RNase-Man 0, obtained by incubating the pure glycoform RNase-Man 1 (prepared as above) with Helix pomatia β1-4 mannosidase.

f. RNase GlcNAc, the product of the digestion of RNase B with Endo H.

g. A co-injection of RNase A with RNase GlcNAc (from f)

h. Electrophoretically pure RNase A.

4.3 Glycosylation alters the global dynamic stability of RNase

The effects of variable glycosylation on the dynamics of RNase was explored by comparing NH proton exchange rates of RNase A and B and the RNase Man5 and Man1 glycoforms (Rudd et al 1994). The exchange of amide protons with solvent (D_2O), which is a measure of dynamic stability or the solvent accessibility and unfolding of the protein, is decreased from 1.5-6 times by glycosylation of the RNase molecule; RNase B is therefore more stable than RNase A. Individual amino acid residues in RNase B affected by the presence of the sugars included His12, His119, Lys41 and Asp121 in the active site, and Arg 39, Arg 85 and Lys31 which form salt bridges to phosphate groups on RNA. Comparisons of exchange rates in different glycoforms showed that protons at residue numbers 10,11,12,32,34 and 35 had relatively slower exchange rates which were dependent on the glycoform type. In the Man-6 to Man-9 glycoforms these protons exchanged faster than in the Man-5 and Man-1 glycoforms. This suggests that the conformation of the Man6-9 oligosaccharides relative to the protein allows greater accessibility to the protein surface in these glycoforms compared with the Man5-Man1 glycoforms.

4.4 Glycosylation protects all glycoforms equally against proteolysis.

In digestions of mixtures of RNase A and each of the Man-5, Man-1, Man-0 or RNaseGlcNAc glycoforms all of the glycosylated enzymes were approximately 20% +/- 3% less susceptible to the broad spectrum protease, Pronase, than RNase A (Rudd et al 1994). The overall rate of cleavage depends on the dynamic stability of the protein which includes the rate at which it unfolds and may also involve steric protection of susceptible sites on the protein by the sugars. The NH-ND exchange experiments demonstrated that the overall flexibility of all the

glycosylated variants was less than for RNase A; this was consistent with the finding that all of the glycoforms tested showed increased resistance to the protease which cleaves ribonuclease at many sites, and with dynamic simulations (Woods et al 1994) which have shown that a large part of the surface of RNase B is shielded by the sugars.

4.5 Preliminary data suggests that RNase oligosaccharides may shield one or more antigenic sites from IgGαRNaseA

Figure 6 shows the results of ELISA assays testing the affinity of a polyclonal IgGαRNase A antibody for some of the glycoforms of RNase B. In figure 6a the wells of a microtitre plate have been coated with mixtures of RNase Man1 and RNase A in proportions ranging from 0-100% of RNase A and 100-0% of RNase Man1. The polyclonal IgGαRNase A antibody binds equally in all the wells, regardless of the relative proportions of RNase A and Man1. This suggests that the antibody is equally able to recognise both molecules and that the Man1 sugar does not shield antibody recognition sites on the protein. In contrast, Figure 6b shows that wells coated with 100% Man5 bind less than 50% of the antibody bound by wells coated with 100% RNase A. Increasing the concentration of RNase A relative to Man5 increases the amount of bound antibody until a plateau is reached when the proportion of RNase A:RNaseMan5 bound to the wells is approximately 40:60%. This indicates that the Man5 glycoform has a lower affinity for the antibody than RNase A. It suggests, as expected, that the polyclonal antibody recognises more than one site on RNaseA and that the sugars attached to the Man 5 glycoform are able to shield one or more, but not all, of the antigenic sites from the antibody. However the interpretation of this data is complex and necessarily qualitative, since there are other factors which may contribute to this effect. These include (a) the relatively large size of the sugars compared with the protein; this may reduce the number of glycosylated molecules bound to the plate compared to RNase A and (b) the orientation of the molecules bound to the plate. While RNase A may be randomly oriented, the glycosylated protein may take up a restricted number of orientations determined by the fact that the sugar moiety cannot bind to the plate. Both these factors would result in a linear increase in binding across the whole range of mixtures of Man5 and RNase A (from 0-100% RNase A), suggesting that they are not the dominant factors in this assay system. Interestingly, the data in 6a suggest that neither the orientation of the RNase Man1 protein nor the number of molecules bound to the plate are significant factors in this case. In figure 6c RNase A has been tested against the mixture of RNase B glycoforms which contains 50% of the Man 5 structure. There is no significant difference between these data and those in Fig.6b, indicating that increasing the size of the oligosaccharides beyond Man 5 does not alter the affinity of the antibody. This is consistent with dynamics simulations of RNase B which indicate that while the Man9 sugars extend further from the protein surface than Man 5 structure, this leads to only a 20% increase in the cross sectional area of the oligosaccharide chain (Woods et al 1994). This suggests that the area of the protein shielded by the Man6-9

glycoforms may not be significantly more than in the Man 5 glycoform. 6d shows the result of a control assay in which RNase A was diluted with BSA from 0-100%. The data indicate that the binding of antibody to RNase A is linear with respect to the amount of RNase A bound to the plate.

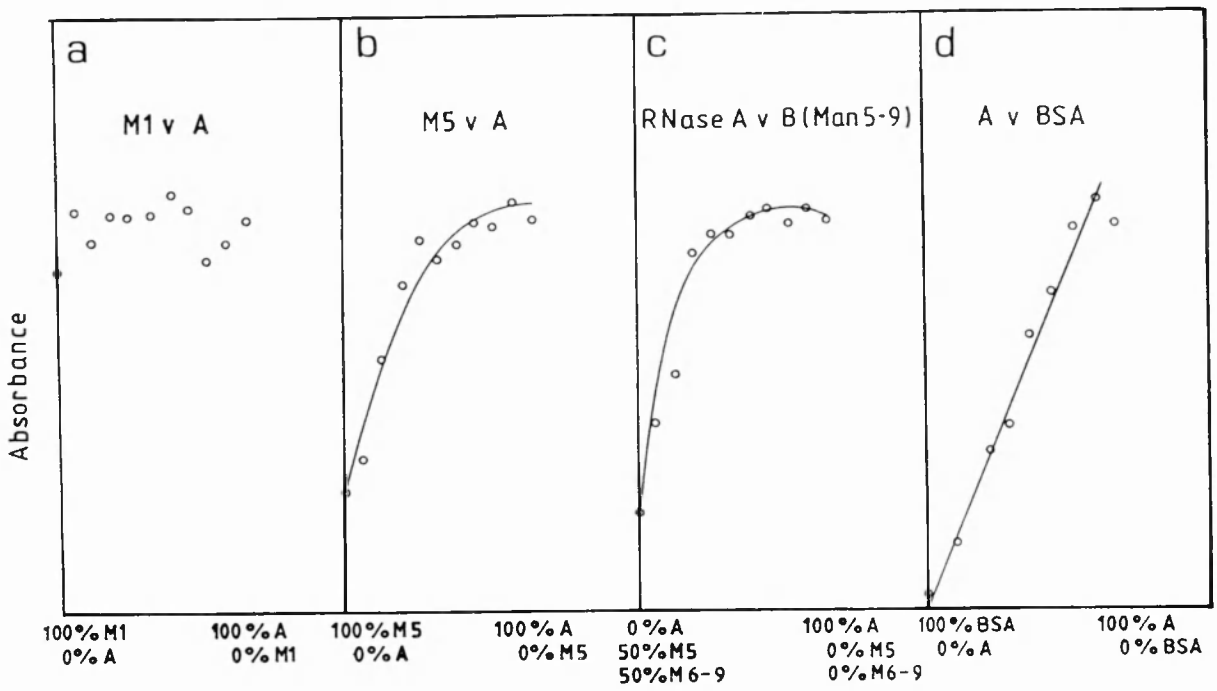


Figure 6: IgG anti RNaseA tested against the Man5 and Man1 glycoforms of RNaseB

Microtitre plate wells coated with RNase B (a) RNaseMan1 (b) RNaseMan5 (c) RNase B and (d) BSA diluted with from 0-100% RNaseA. After washing, polyclonal αIgGRNase A (10ng/ml) was incubated in each plate for 3 hours at 37°C and, after further washing, bound antibody was detected with peroxidase labelled anti-rabbit IgG. The results are the average of eight readings and the error is +/- 10%. The appropriate background (typically 0.08 absorbance units) was subtracted from each point.

4.6 RNase glycoforms show a range of enzymatic activity towards double-stranded RNA

RNase is an endonuclease which cleaves RNA mainly on the 3'-side of pyrimidine nucleotides and many RNA molecules have sections of single stranded RNA interspersed with double stranded sections. The generation of a well defined, *in vitro* labelled double-stranded (ds) RNA substrate enabled the functional activities of RNase A, B, Man-5, Man-1, and Man-0 to be compared in an assay system which monitored the rate of cleavage of ds RNA by each glycoform. The activity of RNase A was more than three times greater than RNase B, and the individual glycoforms were intermediate in activity in the ratios 3.8:2.6:2.2:1.6:1, i.e.

RNase A > Man-0 = Man-1 > Man-5 = RNase B

The naturally occurring mixture of RNase glycoforms showed no significant decrease in activity compared with Man-5 although it contained 50% of glycoforms larger than the Man-5 structure.

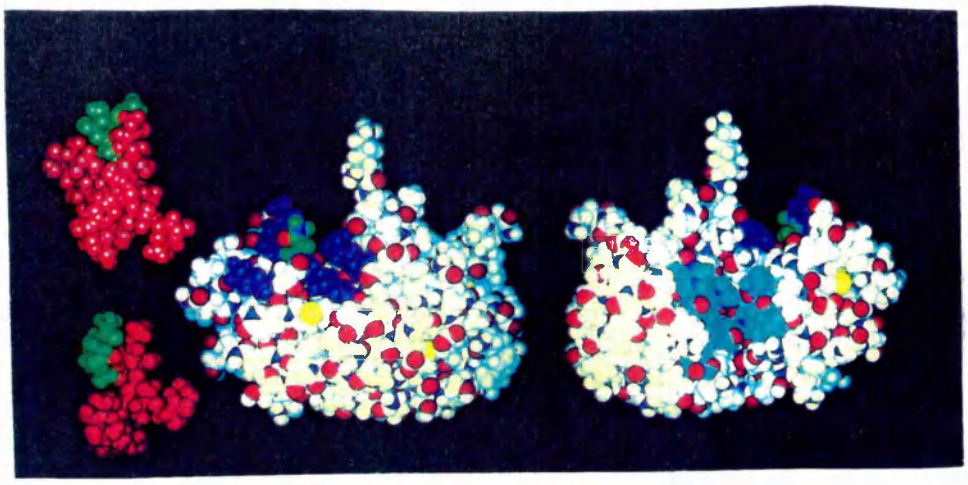


Figure 7: Two views of the unglycosylated RNase molecule with the Man-9 and Man-5 glycoforms drawn to scale

On the protein (white) the active site is cyan, the glycosylation site (Asn34, LeuSer) is green and the cluster of basic residues (Arg10, Lys31, Lys37, Arg33) in range of all the oligosaccharides is dark blue. These residues provide a strong local anion binding site where the 5' terminal phosphate of RNA interacts with the enzyme. Oxygen atoms are red, sulphur atoms yellow and nitrogen atoms blue. Two of the oligosaccharides associated with RNase B are shown on the left of the protein: $GlcNAc_2Man_9$ (upper) and $GlcNAc_2Man_5$ (lower)(Rudd et al 1994). The protein structure is based on the crystal structure (Williams et al 1987).

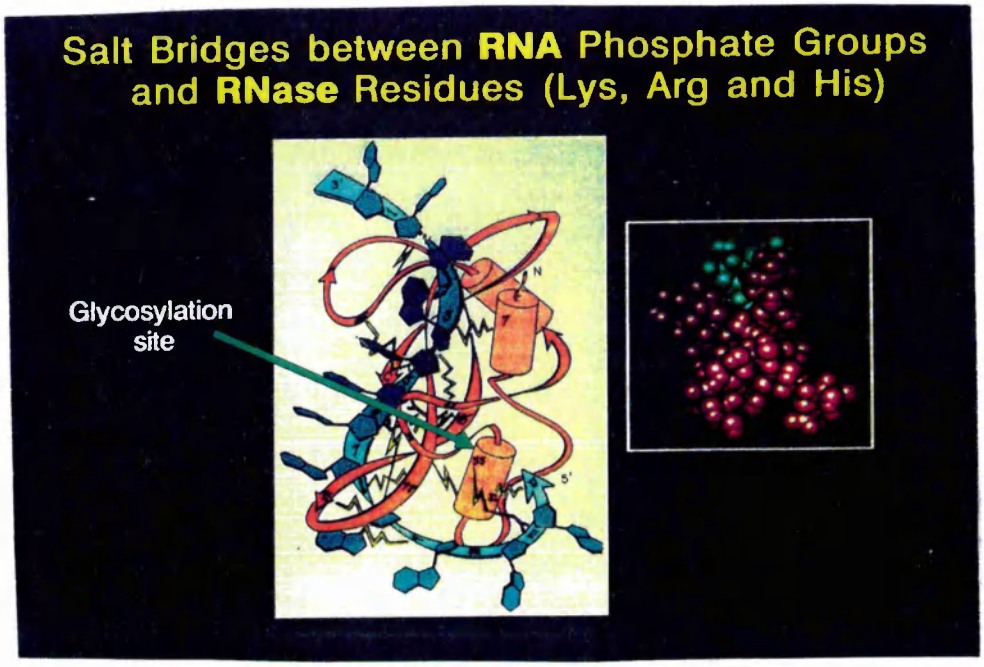


Figure 8: The interaction of RNase B with RNA

A schematic diagram of the interaction of RNase B with RNA showing the anionic binding sites on RNase with which the 5' terminal phosphates of RNA form salt bridges (yellow zigzags). The sugar is shown alongside drawn to scale and Asn34, the site of attachment is marked. (Taken from McPherson et al 1986)

It appears that, while the activity can be increased by removing mannose residues from Man-5, extending the sugar chain beyond Man-5 does not decrease the functional activity of the enzyme further (Rudd et al 1994). These results demonstrate that the glycosylation of RNase is a factor which modulates the rate at which dsRNA is degraded.

4.7 Structure/function relationships probed by molecular modelling

The role of the oligosacccharides in the modulation of the activity of RNase was further explored by molecular modelling. Important amino acid residues for the interaction of RNase with (double-stranded) RNA are those within the active site of the enzyme and also a number which form salt bridges with phosphate groups in the backbone of RNA. There are 5 of these bridges, including one in the active site where cleavage of the phosphodiester bond takes place. In particular, a cluster of residues (Lys 31 Lys 37 Arg 10 and Arg 33) (Fig. 7) provides a strong local anion binding site where the 5' terminal phosphate of RNA interacts with the enzyme (McPherson et al 1986). The size of all the oligosaccharides associated with RNase is such that they can cover this cationic cluster (Fig.8, 9), thus hindering the formation of the salt bridge, and variably reducing the overall fit between the enzyme and substrate.

4.8 Undenatured RNase B is resistant to recombinant PNGase F

Native RNase B was resistant to recombinant PNGase F, indicating that the asparagine-sugar amide linkage is protected by the protein from the active site of this enzyme (Fig.10). In contrast, EndoH is able to cleave the sugar chain to form RNaseGlcNAc (Fig. 5h), indicating that the linkage between the GlcNAc residues in the chitobiose core is accessible.

5. Discussion

5.1 The RNase protein variably protects the oligomannose structures at Asn34 from A.Saitoi α1,2 mannosidase.

The rate of *A.Saitoi* mannosidase digestion of the released oligomannose glycans from RNase B is over 100 times faster than the digestion under identical conditions of the same sugars attached to the protein. Reaction rates depend on several factors; in this study the two major variables were (i) the different diffusion rates of the free sugars and the glycosylated protein and (ii) the steric effects of the protein on the accessibility of the sugars attached to RNase B. To assess whether the observed difference in rate could be accounted for by the difference in diffusion rates between the free and bound sugars, diffusion coefficients (DC) were compared. The DC for RNase A is 1.068×10^{-6} (Van Holde 1971). This may be higher than for RNase B which contains the glycans attached to the protein, together with associated water molecules. The DC for the oligomannose sugars is not available, but would be predicted to be less than the value for the disaccharide sucrose, which is 4.5×10^{-6} (Van Holde 1971). Clearly there is insufficient information to make an accurate quantitative assessment, however the observed difference in rate is more than 100 and it seems unlikely that this can be attributed entirely to differences in diffusion rates. Consistent with this is the observation that the Man6 glycoform is

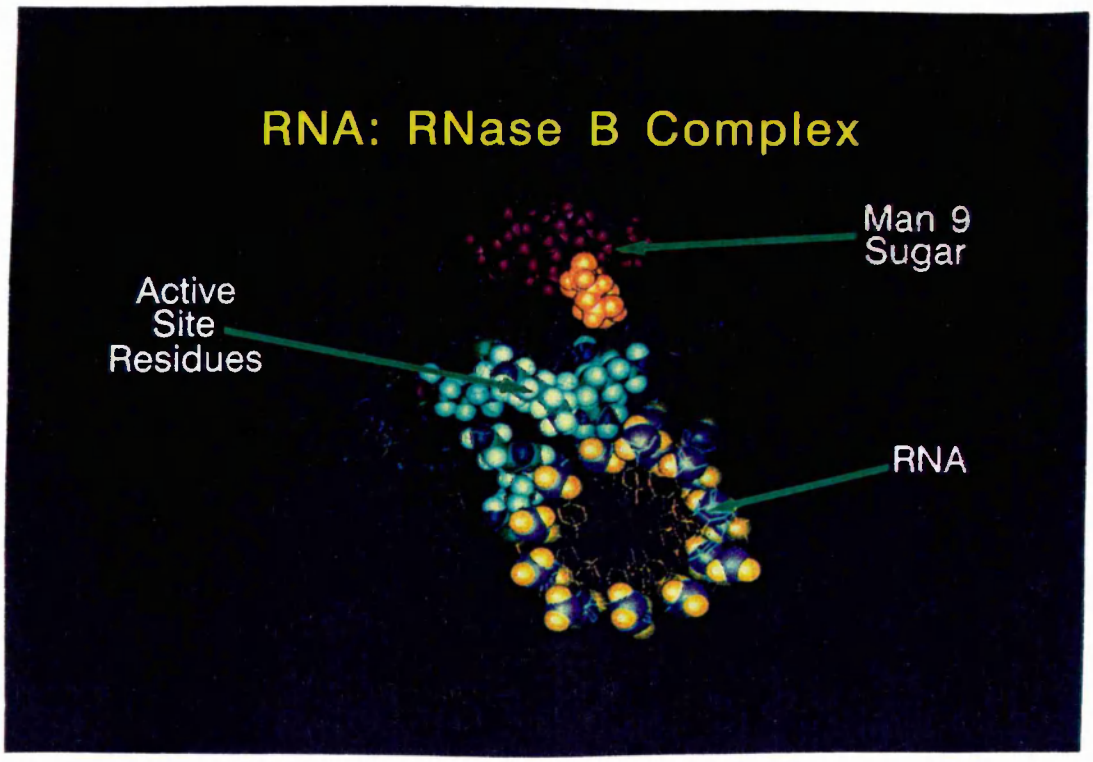


Figure 9: A molecular model of RNase B interacting with RNA.

The active site residues of RNase B are shown in blue, green and white spheres in CPK. The RNA helix is shown with the phosphate groups on a cross section of the spiral in dark blue and yellow spheres in CPK. The phosphate groups are shown interacting with the enzyme through one of the anionic binding sites. the active site. The Man 9 sugar is shown in red CPK with the GlcNAc residue linking the sugar to Asn 34 in yellow. The sugar has been added only to show relative sizes of the oligosaccharide, the enzyme and the section of RNA since the exact orientation of the sugar with respect to the protein is not known. (Rudd, P.M and Saunders, M.R. - unpublished data.)

more resistant to the enzyme than the higher glycoforms. The DC of Man6 would be less than Man7,8 and 9 therefore, if the reaction were diffusion controlled, the rate for Man6 would be faster than that for the higher glycoforms, not slower as was observed. Since each glycoform is a different substrate for the enzyme, every digestion will have different kinetic characteristics. These data suggest that the slower rate of digestion of the α1,2 terminal mannose sugars on Man6 is a result of its lower affinity for the enzyme and that this may be due to its proximity to the protein matrix. This indicates that in RNase B the protein structure may play a role in protecting the sugars from *A.Saitoi* mannosidase. The biosynthetic processes involve the conversion of Man9-6 to the Man5 glycoform (see Ch.1). The natural mannosidase is bovine pancreatic mannosidase not *A.Saitoi* mannosidase; therefore it is not justified to extrapolate to the *in vivo* system, however these data do not preclude the possibility that the protein may give some protection to the oligomannose sugars in the native RNase B, reducing the efficiency of the enzyme so that only 50% of the natural mixture of glycoforms is converted to RNaseMan 5.

5.2 Oligomannose sugars shield part of the surface of RNase B and modulate the activity of the enzyme

The effect of variable glycosylation on the environment of important regions of the protein may be one of the factors by which the activity of RNase is modulated. The extent to which the oligosaccharides of RNase B can shield the protein surface has been examined in a computer simulation of the dynamic properties of the Man 9 sugars (Woods et al 1994, Fig.1). The simulation indicated that the core residues (Manβ1-4GlcNAcβ1,4GlcNAc) maintain a relatively stable conformation. For other glycoproteins, such as *Erythrina corallodandron* lectin (Shaanan et al 1991), where X-ray crystal structures clearly define the N-linked core residues, the oligosaccharides vary in sequence yet the core conformations are remarkably similar. This finding allowed the computer simulated model of the vibrating Man9 sugars to be linked to the X-ray structure of RNaseB. Variation of the side chain angles of the GlcNAc-Asn linkage over +/-30° showed that this motion had a major effect on the volume of space occupied by the oligosaccharide and consequently on the area of the protein surface shielded by the sugar. Interestingly, the modelling also suggests that Man5 may shield 80% of the area covered by Man9. This comparatively small difference is partly a result of the tight hydrophobic stacking of the terminal α1,2 linkages in Man9 (D1,D2,D3) against the penultimate mannose residues A,B,C (Fig. 2b). The result is that there is only a 20% increase in the cross sectional area of the Man9 oligosaccharide over the Man5 structure. This predicts that, if steric interactions between RNA and the oligosaccharides contribute to the decrease in enzyme activity, increasing the oligosaccharide size up to Man5 would lead to decreased activity while Man5 to Man9 glycoforms would cause no further decrease in the reaction rate. In this thesis the naturally occurring mixture of RNase glycoforms showed no significant decrease in activity with RNA compared with Man-5 although it contained 50% of glycoforms larger than the Man-5 structure. It appears that, while the activity can be increased by removing mannose residues from Man-5,

extending the sugar chain beyond Man-5 does not decrease the functional activity of the enzyme further. (RNase A (3.8) >Man0 (2.6) = Man1 (2.2) >Man5 (1.4) =RNaseB (1))
Additional molecular modelling studies in this thesis suggest that different sized oligosaccharides may exercise variable steric effects both at the active site and at a cluster of residues involved in the formation of salt bridges to phosphate in RNA (Fig.7) decreasing the stability of the RNase/RNA complex (Fig. 8) and the rate of cleavage.

5.3 RNase glycoforms interact variably with α-RNase A

Probing the surface of glycosylated RNase with an antibody raised against the unglycosylated molecule suggested that the Man5-9 sugars protect at least one antigenic epitope while the Man1 sugar does not. The approximate size of an antigenic epitope is between 5-7 amino acids (roughly 7x12x35A)

5.4 The dynamic stability of RNase B glycoforms is greater than RNase A

The dynamic stability of RNase B and the Man5 and Man1 glycoforms was examined in NH-ND exchange experiments (Joao in Rudd et al 1994).The results indicated that the overall flexibility of all the glycosylated variants was less than for RNase A. This overall increase in rigidity may contribute to the increased resistance to Pronase and the decreased functional activity of the glycosylated variants compared with RNase A. Differences in the dynamic stability of the same amino acid residues in different glycoforms, including ones at the active site and an anionic binding site, may contribute to the > 3-fold variation in the functional activity of the glycoforms.

5.5 Glycosylation protects Man 0,1 and 5 glycoforms equally against proteolysis

The dynamics simulations described in 5.2 above indicate that a substantial part of the protein surface is shielded by the Man5-9 oligomannoses structures. However shielding by Man1, Man0 or a single GlcNAc residue is less significant suggesting that other factors may be important in the finding that all of the glycoforms showed increased resistance to Pronase. Pronase is a broad spectrum protease which cleaves ribonuclease at many sites. In digestions of mixtures of RNase A and each of the Man-5, Man-1, Man-0 or RNaseGlcNAc glycoforms the glycosylated enzymes were equally less susceptible (by 20%) to digestion by the protease than RNase A. One factor which controls the overall rate of cleavage is the dynamic stability of the protein which includes the rate at which it unfolds. The dynamic stability of RNase B was examined in NH-ND exchange experiments and the results indicated that the overall flexibility of all the glycosylated variants was less than for RNase A. This overall increase in rigidity may contribute to the increased protease resistance of the glycosylated variants compared with RNase A.

5.6 The protein structure of RNase B limits access to the glycans and may influence processing

The role of the RNase protein in protecting the sugar has been probed with *A.Saitoi* mannosidase and the data indicate that the α1,2 terminal mannose residue on the Man 6

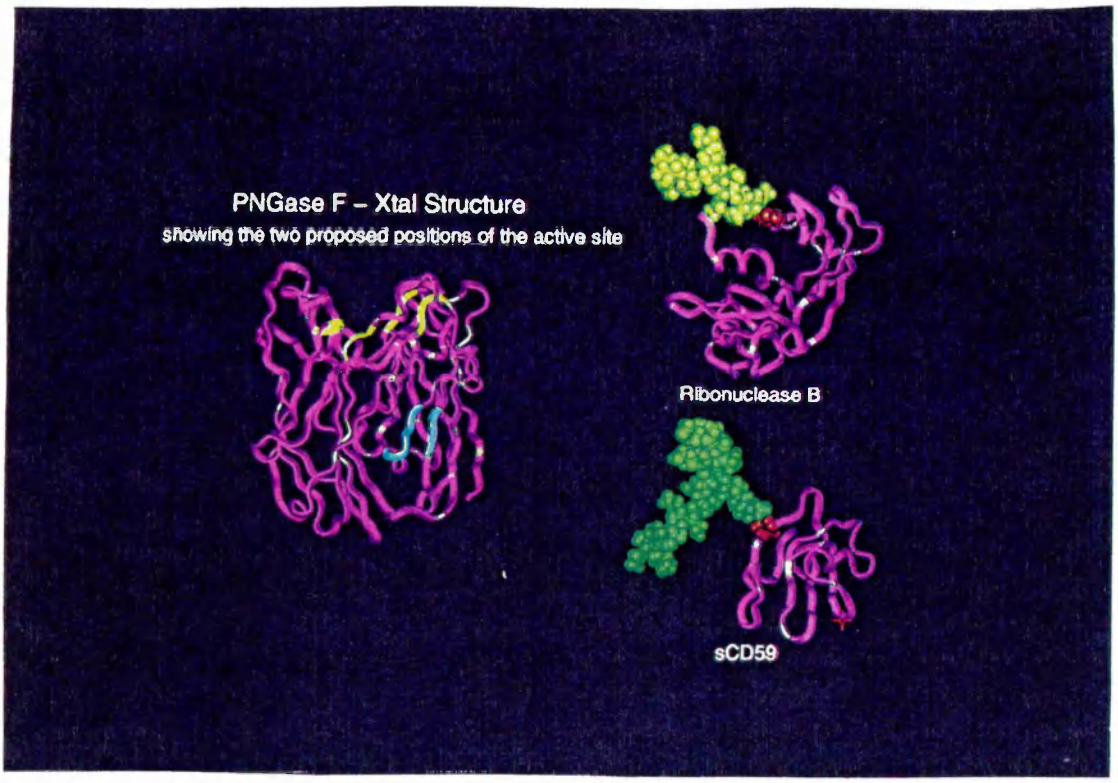


Figure 10: The crystal structures of RNase B, CD59 and PNGase F

The structure of PNGase F is based on the crystal structure of Norris et al 1994. The protein is folded into two domains, each with an eight stranded antiparallel β-jelly roll configuration similar to the types of structures found in lectins. RNase B was modelled according to the crystal structure of Williams et al 1987. CD59 was modelled according to the NMR solution structure of Kieffer et al 1994. All three molecules are modelled to the same scale. Two active sites have been proposed for PNGase F: (i) Thr42, Lys44, Asp99 and Thr101 on strands D and G of domain 1 (ii) a site in a cleft at the opposite end of the molecule, flanked by five tryptophan residues and one phenylalanine residue. In contrast to the CD59 asparagine-sugar amide linkage, which is susceptible to the enzyme, the same linkage in RNase is protected by the protein and not accessible.

glycoform is more resistant to cleavage than the terminal α1,2 oligosaccharides on the larger glycoforms. The protein limits access to the sugar and while endo H can cleave the oligomannose structures attached to the native glycoprotein within the chitobiose core, PNGase F is not able to cleave the asparagine - sugar amide linkage. The physical accessibility of oligosaccharides to the processing enzymes, as determined by the folding of the polypeptide chain, can influence the extent of oligosaccharide processing. Accessibility has been probed with enzymes such as endoglycosidase H, which cleave the oligosaccharide chain close to the site of attachment to the protein. Endo H has been used to examine the oligosaccharide chains attached to Sindbis glycoproteins. For Sindbis virus grown in CHO cells deficient in N-acetylglucosaminyl transferase Endo H (which cleaves between the GlcNAc residues in the chitobiose core) preferentially cleaved the unprocessed oligomannose glycans located at sites that normally contain complex sugars (Hsieh et al 1983). It has therefore been suggested that the inaccessibility of this linkage to Endo H is an indicator that structures will remain unprocessed. However, contrary to this, in this thesis endo H was shown to release the oligomannose glycans from native RNase B. In addition, RNase B from other organs, such as spleen, is glycosylated with complex as well as oligomannose sugars (Lawrence et al 1983 a,b, Yamashita et al 1986). Interestingly, it has been shown by Williams and Lennarz (1984) that RNase expressed in the presence of a reducing agent which prevents the formation of disulphide bonds, contains fully processed oligosaccharides. A similar experiment involving tissue plasminogen activator showed that prevention of protein folding resulted in 100% occupancy of Asn 184, a site which is naturally variably occupied (Allen et al 1995). These results indicate that the conformation of the protein may influence both the processing of glycans and the site occupancy.

In common with tissue plasminogen activator and plasminogen, RNase B has a variably occupied N-glycosylation site. The recent finding (Suzuki et al 1993) that several mouse and human cell lines [L-929 (mouse), BALB-3T3 (mouse fibroblast), BALB/c mouse plastocytoma and human skin fibroblast] possess natural PNGase activity suggests that the specific removal of N-linked glycans *in vivo* may be an important biosynthetic mechanism. For example, in plasminogen type II (ch.7) the unoccupied glycosylation site contains aspartic acid not asparagine, suggesting that this glycoform is the result of the removal of the N-linked sugar present in type I plasminogen (S. Innoue - personal communication). The protection of the asparagine-sugar linkage from inappropriate deglycosylation by PNGase may therefore be an important function of the local protein conformation. In the case of RNase B the protein structure protects the Asn - sugar amide linkage from one particular peptide-N-glycosidase F (F. meningosepticum) (Fig.10).

5.7 Conclusion

Pure individual glycoforms of bovine pancreatic RNase B have been prepared by enzymatic modifications of the native glycoprotein and analysed by CE. In a series of studies of native RNase B, the individual glycoforms and RNase A it was found (i) the protein variably protects

the oligomannose sugars from exoglycosidase digestion (ii) the protein protects the sugar-amide linkage, but not the glycosidic linkage within the chitobiose core from PNGase F and EndoH respectively (iii) glycosylation increases the protease resistance of RNase B independent of glycoform type (iv) the individual glycoforms were used to demonstrate that the size of the attached oligosaccharide is a factor in protecting antigenic sites and in modulating the interaction between enzyme and substrate (v) the individual glycoforms were used to demonstrate that there are increases in dynamic stability throughout the whole molecule including active site residues and residues involved in the formation of salt bridges between RNase and RNA.

In principle such dynamic and steric effects of glycosylation may be expected to apply to all glycoproteins. The exact nature of the glycoform may also be an important feature of control for enzymes (such as tissue plasminogen activator discussed in ch.7) which are multidomain and whose interactions involve molecular rearrangements.

7. References to chapter 4:

D'Alessio, G., Di Donato, A., Parente, A., & Piccoli, R. (1991) Trends Biochem. Sci., 16 104-106 Seminal RNase: a unique member of the ribonuclease superfamily.

Allen, S., Naim, H.Y. and Bullied, N.J. (1995) J. Biol. Chem. 270 4797-4804 Intracellular folding of tissue type plasminogen activator; effects of disulphide bond formation on N-linked glycosylation and secretion.

Fisher, A.J. & Johnson, J.E. (1993) Nature 361 176-9 Ordered duplex RNA controls capsid architecture in an icosahedral animal virus.

Hsieh, P., Rosner, M.R. and Robbins P.W. (1983) J. Biol. Chem. 258 2555-2561 Selective cleavage by Endo β-N-acetyl glucosaminidase H at individual glycosylation sites of Sindbis virus envelope glycoproteins.

Joao, H.C., Scragg, I.G., & Dwek, R.A. (1992) FEBS Lett. 307, 343-346 Effect of glycosylation on protein conformation and amide proton exchange rates in RNase B.

Lawrence, C.W., Little, P.A., Little, B.W., Miller, M.J., Bazel, S. & Alhadeff, J.A. (1993a) Glycobiology 3 241-248 Human non-secretory ribonucleases. I. Purification, peptide mapping and lectin blotting analysis of the kidney, liver and spleen enzymes.

Lawrence, C.W., Little, P.A., Little, B.W., Glushka, J., & van Halbeek, H. & Alhadeff, J.A. (1993b) Glycobiology 3 249-259 Human non-secretory ribonucleases. II. Structural characterisation of the N-glycans of the kidney, liver and spleen enzymes by NMR spectroscopy and electrospray mass spectrometry.

McPherson, G., Brayer, D., Cascio, R., and Williams, R. (1986) Science 232, 765-768 The mechanism of binding of a polynucleotide chain to pancreatic ribonuclease.

Mori, K., Dwek, R.A., Downing, A.K., Opdenakker, G., and Rudd, P.M. (1995) J. Biol. Chem. 270 3261-3267 The activation of type 1 and type 2 plasminogen by type I and type II tissue plasminogen activator

Norris, G.E., Stillman, T.J., Anderson, B.F. and Baker, E.N. Current Biology Structure (1994) 2:1049-1059 The three dimensional structure of PNGase F, a glycosyl-asparaginase from *Flavobacterium meningosepticum.*

Parekh, R.B., Dwek, R.A., Sutton, B.J., Fernandes, D.L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, K., Takeuchi, f., Nagano, Y.,

Miyamoto, T. and Kobata, A. (1985) Nature 316, 452-457 Association of Rheumatoid Arthritis and Primary Osteoarthritis with changes in the glycosylation pattern of total serum IgG.

Puett, D. (1973) J. Biol. Chem. 248 3566-3572 Conformational studies on a glycosylated bovine pancreatic ribonuclease.

Rademacher, T.W., Parekh, R.B., & Dwek, R.A. (1988) Ann. Rev. Biochem. 57, 785-838 Analysis of glycoprotein associated oligosaccharides.

Reddi, K. and Holland, J.F. (1976) Proc. Natl. Acad. Sci. USA 73 2308-2310 Elevated serum ribonuclease in patients with pancreatic cancer.

Rudd, P.M., Scragg, I.G., Coghill, E. & Dwek, R.A. (1992) Glycoconjugate Journal 9, 86-91 The Separation and Analysis of the Glycoform Populations of Ribonuclease B using Capillary Electrophoresis.

Rudd, P.M., Joao, H.C., Coghill, E., Fiten, P., Saunders,M. R. Opdenakker, G. and Dwek, R.A. (1994) Biochemistry 33, 17-22 Glycoforms modify the dynamic stability and functional activity of an enzyme.

Shall, S. & Barnard, E.A. (1969) J. Mol. Biol. 41 237-251 Thiolation of ribonuclease.

Shaanan, B., Lis, H. and Sharon, N. (1991) Science 254 862-866 Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose.

Stevens, J.G., Wagner, E.K., Devi-Rao, G.B., Cook, M.L. & Feldman, L.T. (1987) Science 235 1056-9 RNA complementary to a Herpes virus a gene mRNA is prominent in latently infected neurons.

Suziki, T., Seko, A., Kitajima, K., Inoue, Y. and Inoue, S. (1993) BBRC 194 1124-1130 Identification of peptide: N-glycanase activity in mammalian-derived cultured cells.

Van Damme, J., Opdenakker, G., Billiau, A., de Somer, P., de Wit, L., Poupart, P. & Content, J. (1985) J. Gen. Virol. 66, 693-700 Simulation of fibroblast interferon production by a 22k protein from human leukocytes.

Van Damme, J., Cayphas, S., Opdenakker, G., Billiau, A., Van Snick, J. (1987) Eur. J. Immunol 17 1-7 Interleukin 1 and poly (rI)-poly (rC) induce production of a hybridoma growth factor by human fibroblasts.

Van Damme, J., Decock, B., Bertini, R., Conings, R., Lenaerts, J-P., Put, W., Opdenakker, G., and Mantovani, A. (1991) Eur. J. Biochem. 199 223-229 Production and identification of natural monocyte chemotactic protein from virally infected murine fibroblasts.

Van Holde, K.E. (1971) Physical Biochemistry p.89 Published by Prentice-Hall Inc., Englewood Cliffs, N.J. USA

Williams, R.L. & Greene, S.M., & McPherson, A. (1987) J. Biol. Chem. 263 16020-16031 The crystal structure of ribonuclease B at 2.5A resolution.

Williams, D.B. and Lennarz, W.J. (1984) J. Biol. Chem. 259 5105-5114 Control of asparagine-linked oligosaccharide chain processing: studies on bovine pancreatic ribonuclease B.

Wittwer, A.J., Howard, S.C., Carr, L.S., Harakas, N.K., and Feder, J., Parekh, R.B., Rudd, P.M., Dwek, R.A., and Rademacher, T.W. (1989) Biochemistry 28 7662-7669 Effects of N-glycosylation on *in vitro* activity of Bowes Melanoma and human colon fibroblast derived tissue plasminogen activator.

Wlodawer, A., Svensson, A.L., Sjolin, L., & Gilliland, G.L. (1988) Biochemistry 27, 2705 Structure of phosphate-free ribonuclease A refined at 1.26A.

Woods, R.J., Edge, C.J., and Dwek, R.A. (1994) Nature Structural Biology 1 499-501 Protein surface oligosaccharides and protein function.

Wool, I.G. (1984) Trends Biochem. Science 9 14-17 The mechanism of action of the cytotoxic nuclease α-sarcin and its use to analyse ribosome structure.

Yamashita, K., Hitoi, A., Irie, M., & Kobata, A. (1986) Communication Arch. Biochem. and Biophysics 250 263-266 Fractionation by lectin affinity chromatography indicates that the glycosylation of most ribonucleases in humans viscera and body fluids is organ specific.

Chapter 5

The Identification of Abnormal Glycoform Populations in Transferrin from Carbohydrate Deficient Glycoprotein Syndrome Serum by Capillary Electrophoresis may lead to a rapid diagnostic test.

Publications associated with this chapter:

(i) Rudd, P.M., Scragg, I.G., Coghill, E. & Dwek, R.A. (1992) Glycoconjugate Journal 9, 86-91 Separation and analysis of the glycoform populations of RNase B using capillary electrophoresis.

(ii) Rudd, P.M., Mattu, T. and Honda, S. in 'Capillary Electrophoresis: A Practical Approach' (OUP) Ed. Goodall, D. (In press) 'Capillary Electrophoresis of Oligosaccharides, Glycoproteins and Glycopeptides'.

Acknowledgements:

1. This study was undertaken in collaboration with Dr. Brian Winchester and Dr. Nasi Mian, Great Ormond Street Hospital, London, who provided the samples of transferrin.
2. The work was carried out in conjunction with Mr. Taj Mattu and Mr. Oleg Iourin (The Glycobiology Institute, Oxford) who performed the capillary electrophoresis.

Abbreviations: CDGS: Carbohydrate Deficient Glycoprotein Syndrome; CE: Capillary Electrophoresis; CZE: Capillary Zone Electrophoresis (free solution CE); IEF: Iso-electric focusing; RNase: ribonuclease; S0-S6: Subsets of glycoform populations of human transferrin which contain from 0-6 sialic acid residues attached to complex glycans; J: non-glycosylated transferrin; endo D: endoglycosidase D; GlcNAc: N-acetyl glucosamine; NHT: normal human transferrin; SA2H5N4: glycan composition formula of oligosaccharide containing 2xsialic acid, 5xhexose and 4xN-acetyl hexosamine glycans, ER: endoplasmic reticulum.

Abstract

One of the characteristics of Carbohydrate Deficient Glycoprotein Syndrome (CDGS) is the presence in serum of abnormal glycoforms of transferrin. While the amino acid sequence and the glycans attached to CDGS transferrin are identical to those associated with the normal glycoprotein, the abnormal transferrin contains a higher proportion of unoccupied N-linked glycosylation sites. In transferrin from normal serum, over 95% of the two glycosylation sites are invariably occupied with either the bi- or tri-antennary non-fucosylated, fully sialylated complex glycans, SA2H5N4 and SA3H6N5. Both glycoform heterogeneity and variable site occupancy may lead to the generation of a range of glycoforms which contain from 6-0 sialic acid residues and therefore differ in the amount of negative charge they contain. Capillary electrophoresis (CE) has been used to resolve quantitatively the intact glycoform populations of both normal and CDGS transferrin at the protein level. The separation was based on the charge differences between the glycoform populations and involved using a buffer containing sodium glutamate. The individual glycoforms were identified using CE to monitor sequential exoglycosidase digestions of the sugars on the intact glycoprotein and the elution positions of the different glycoforms were determined. Abnormal levels of glycoforms containing one unoccupied glycosylation site and non-glycosylated transferrin, neither of which has been detected in normal transferrin, were identified in transferrin from CDGS patients. A rapid diagnostic test for the disease, which may also be an indicator of disease severity, has therefore been proposed which allows the glycoform populations of CDGS transferrin to be compared quantitatively with those in normal transferrin.

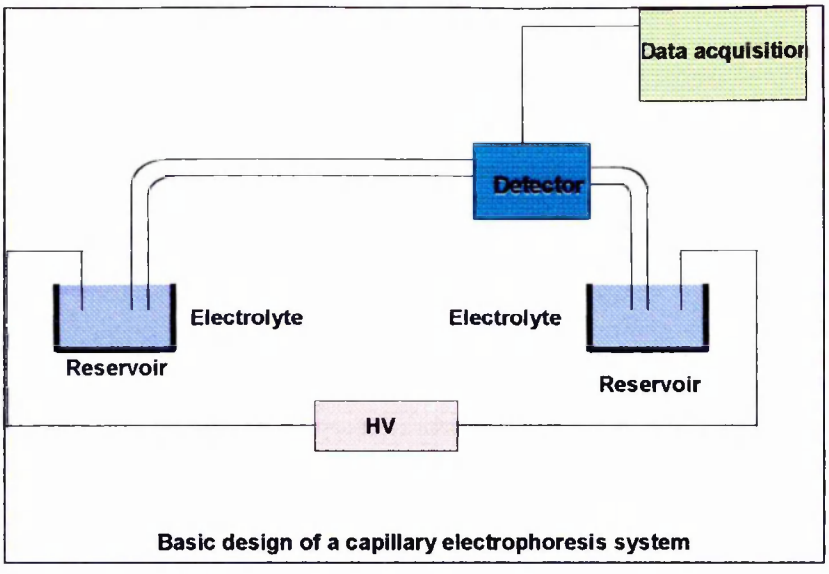


Figure 1a: The basic CE instrument consists of a fused silica capillary with an optical window, a UV detector, a high voltage power supply, two electrode assemblies and two buffer reservoirs. Field strengths higher than 500 volts/cm can be applied to the narrow capillary allowing highly enhanced resolution compared with conventional gel electrophoresis.

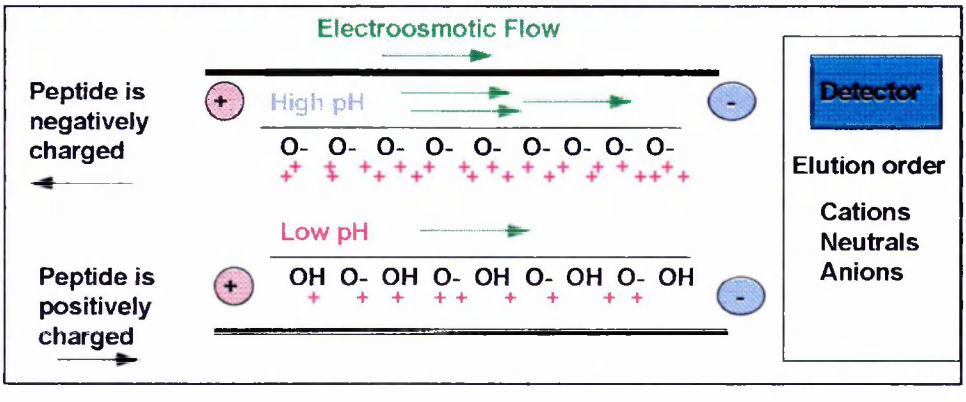


Figure 1b: Effect of electroosmotic flow (EOF). In uncoated capillaries EOF is always towards the cathode. At high pH the capillary walls are strongly negatively charged and the EOF is greater than at low pH. At high pH most peptides are negatively charged and migrate towards the anode, however EOF is always stronger and anions are eventually swept to the cathode. At low pH most peptides are positively charged and move towards the cathode with the EOF. The elution order is the same in both cases: cations elute first, then neutrals and finally anions.

The Identification of Abnormal Glycoform Populations in Transferrin from Carbohydrate Deficient Glycoprotein Syndrome Serum by Capillary Electrophoresis may lead to a rapid diagnostic test.

1. The structure of glycoforms

1.1 The assignment of glycoform structure by analysis of released oligosaccharides

Many current techniques for determining glycoform populations involve the analysis of oligosaccharide libraries obtained by releasing the sugars either from the intact glycoprotein or from glycopeptides. However, for a protein with more than one glycosylation site, the information relating to specific sets of glycoforms is lost, even if the glycosylation at each site can be fully defined. Glycoforms may be assigned from the analysis of the released sugars by assuming that the glycosylation sites on the intact glycoprotein can accommodate all the possible permutations and combinations of the sugars associated with the molecule, but there are many instances where this assumption is incorrect . For example, both CD2 (ch.2) and tPA (ch.7) contain one site which is almost entirely glycosylated with oligomannose structures. Even if a site analysis is carried out there may be constraints on the particular combinations of sugars which can be processed on any one molecule. For example, in Bowes Melanoma tPA the glycosylation analysis of the structures at Asn 448 shows that the site contains more neutral and fewer sulphated glycans in type I tPA when the site at Asn 184 is occupied compared with type II tPA when it is not (Dr. D. Wing - unpublished data. See ch.7 fig. 23). This demonstrates that glycosylation at one site can affect the processing at another and, moreover, it is not known whether specific glycans at site 184 are required to pair with either the sulphated or neutral structures at 448.

1.2 The resolution of glycoforms at the protein level by Capillary Zone Electrophoresis (CZE)

Capillary electrophoresis (CE), which offers the possibility of separating intact glycoforms at the protein level, may therefore be a powerful addition to the existing carbohydrate technologies. At present other electrophoretic techniques such as SDS PAGE and isoelectric focusing (IEF) generate only a limited amount of analytical information relating to intact glycoforms. This is in part because the resolution which can be achieved in these techniques is restricted by the problems of heat dissipation which arise when high voltages are applied to conventional gel systems. In CE, the narrow capillary allows heat to be dissipated rapidly and field strengths higher than 500 volts/cm (up to 30kV) can be applied giving the potential for highly enhanced resolution. The basic CE instrument consists of a fused silica capillary with an optical window, a UV detector, a high voltage power supply, two electrode assemblies and two buffer reservoirs (Fig.1a). In most assay systems the inner wall of the capillary column carries a negative charge because the silanol groups ionise above pH3. Positive ions in the electrolyte are attracted to the wall and, under high voltage, migrate towards the cathode creating a bulk flow of liquid called the electroosmotic flow (EOF) (Fig. 1b). If a sample is introduced into the capillary at the

anode side all the components are carried by the EOF towards the cathode. Simultaneously the current flow will retard the negatively charged species so that the positive species elute first followed by the neutral and finally the negative. An inherent difficulty associated with analysing glycoforms by CE resides in the fact that the homogeneity of the polypeptide chain frequently masks the diversity of the glycans associated with it, so to achieve separation the structural differences between the glycans has to be amplified by complexing with an ionic species. The chemistry which has been used to achieve separations has, for the most part, exploited subtle differences in the conformation of the hydroxyl groups associated with the monosaccharide residues which make up the oligosaccharides. For example, the oligomannose glycoforms of bovine pancreatic RNase B have been separated into their individual glycoforms in their correct molar proportions by CE (Rudd et al 1992). This single domain, globular protein, which is the subject of chapter 4, is a mixture of unglycosylated RNase A (13,690 kd) and a collection of glycoforms (designated RNase B Man-5,-6,-7,-8 and -9) in which the oligomannose series, Man-5 to -9 is associated with the single N-glycosylation site at Asn34LeuSer. RNase Man-1 to -9 glycoforms, generated by enzymatic modifications of the natural population, were separated using borate buffer. Borate ions complex with cis diols and can therefore add a charge to neutral oligosaccharides. The number of charges depends on the number and accessibility of hydroxyl groups in the correct orientation and this may varies from one oligosaccharide to another, giving a basis for an electrophoretic separation. The glycoform ladder showing the resolution of RNase B is shown in figure 2b, and, for comparison, a conventional SDS PAGE analysis of the natural population of RNase B (Man9-5) is shown in figure 2a.

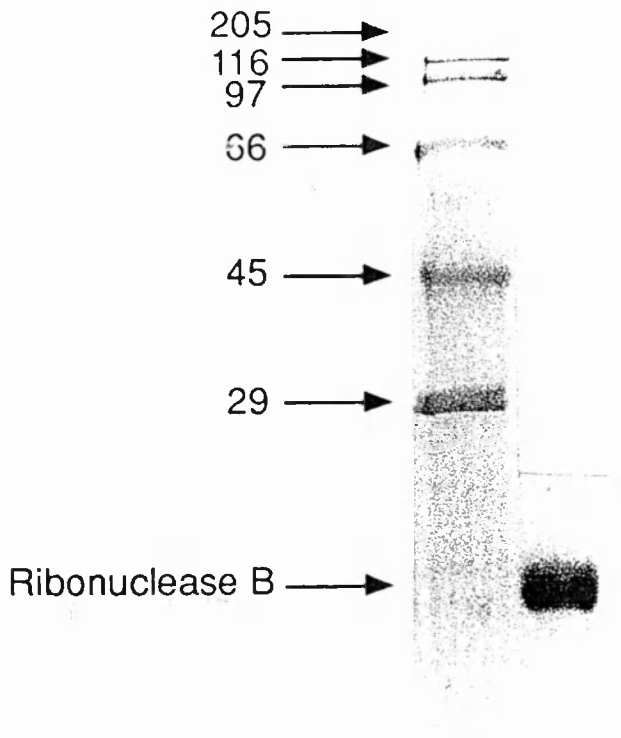


Figure 2: (a) SDS PAGE (15% reducing gel) separation of the natural population of RNase B glycoforms (RNase B Man5-9).

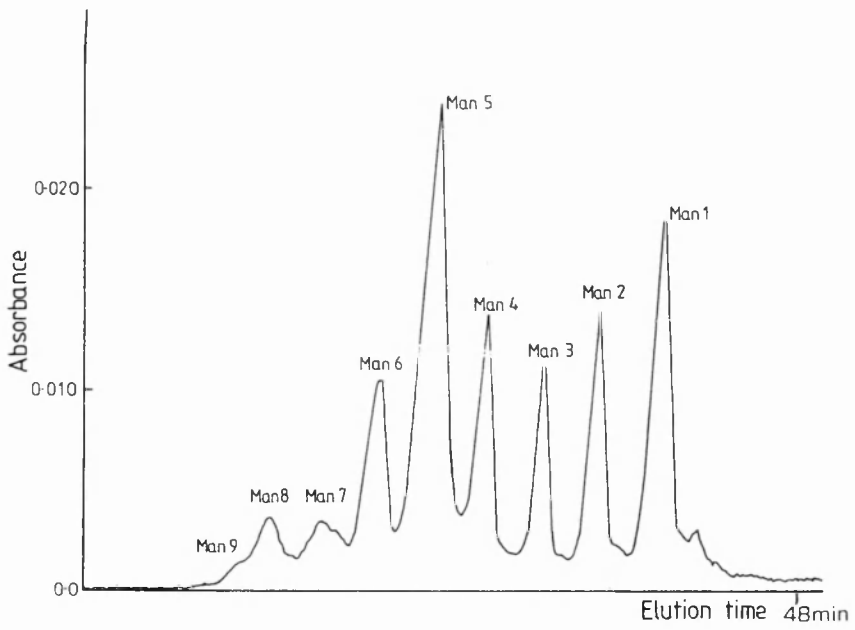


Figure 2 (b) Capillary electrophoresis separation of the set of glycoforms of RNase B containing from 1 to 9 oligomannose residues. *The buffer was 20mM sodium phosphate, 50mM SDS, 5mM sodium tetra borate, pH 7.2.*

To date there have been relatively few studies of glycoproteins or glycopeptides using CE, but those which have been described suggest that there are potentially many ways in which the structure and function of glycoproteins can be explored by this technology. There are currently very few methods available for separating glycoforms at the protein level and in this respect CE has proved to be a valuable addition to the existing technologies; however the methods are not generally applicable, and optimal buffer systems need to be developed for every glycoprotein individually. Once resolution conditions are established the way is open to monitor glycoform alterations in fine detail, and in the case of RNase B, this enabled single pure glycoforms of a glycoprotein to be prepared for the first time (Rudd et al 1992). Successful separations of other glycoproteins include the analysis of tPA glycopeptides (Taverna et al 1992) and erythropoeitin (Tran et al 1991). Within the CE programme in the Glycobiology Institute glycoproteins containing oligomannose sugars (RNase B and prolyl hydroxylase) and neutral complex glycans (ovalbumin) have been resolved, and the work in this chapter describes a complementary project which involved an exploration of the possible application of CE to the analysis of transferrin, a glycoprotein which naturally contains almost exclusively sialylated complex glycans. The technology was developed within a collaborative project (with Dr. Brian Winchester), the aim of which was to develop a rapid method to screen transferrin quantitatively for unglycosylated and monoglycosylated glycoforms characteristic of

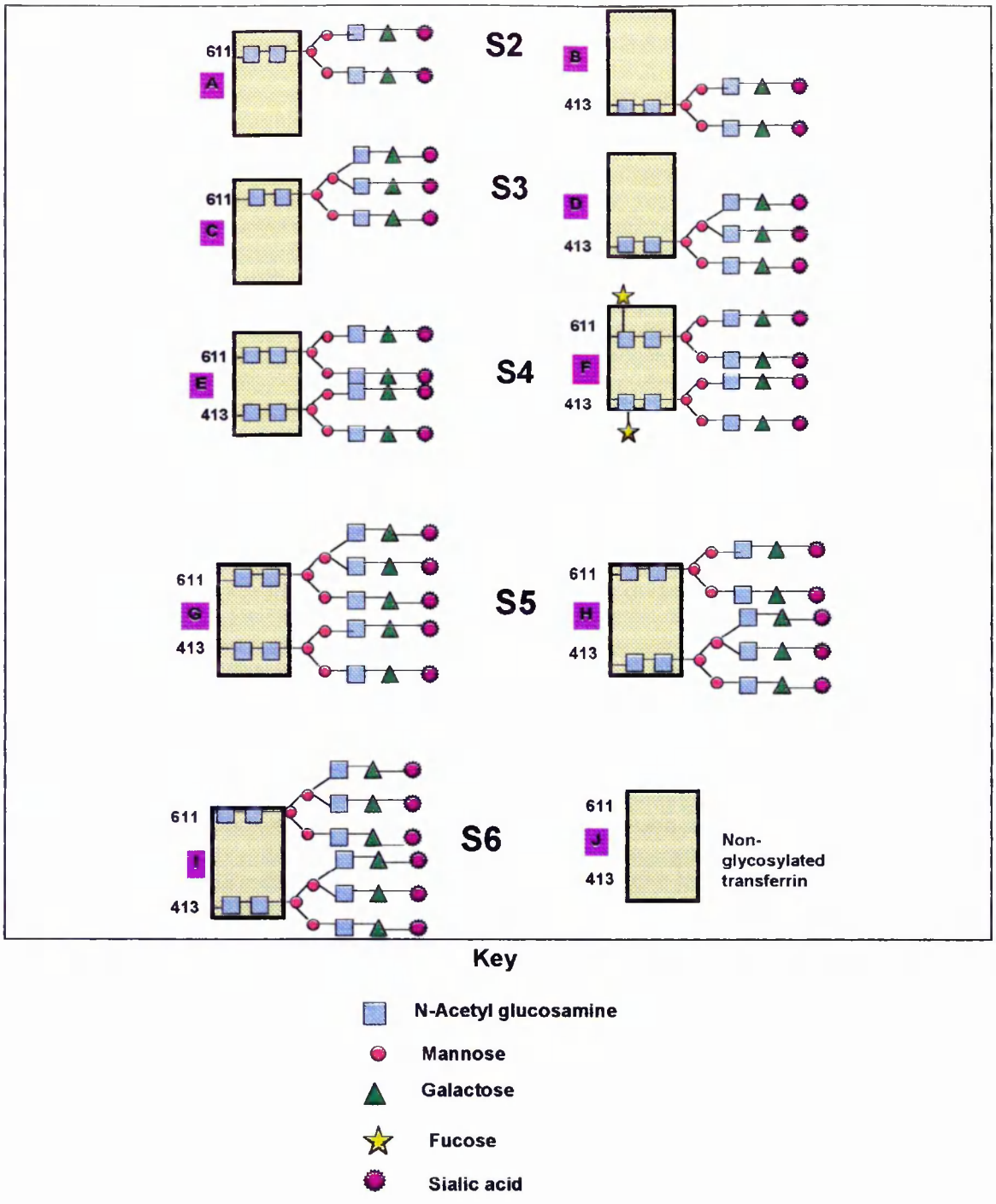


Figure 3: The structures of the set of transferrin glycoforms associated with CDGS.

Transferrin from CDGS patient typically contains up to 37% of glycoforms in which only one glycosylation site is occupied, and 13% in which neither site has been glycosylated (Yamashita et al 1992), although the sugars at the sites are indistinguishable in type and charge from the normal glycans. CDGS transferrin may contain fully sialylated glycoforms with 2,3,4,5 and 6 sialic acid residues (S1-S6) as well as a small population containing A1 and A3 sugars. In addition there is a population of non-glycosylated molecules, designated J. Normal transferrin also contains structures with S2-S6 sialic acid residues, but these are all derived from transferrin with both glycosylation sites occupied.

Carbohydrate Deficient Glycoprotein Syndrome (CDGS). CDGS is characterised biochemically by the absence of glycosylation at normally occupied sites on a wide range of proteins and the current diagnostic procedure is to examine a diagnostic marker, transferrin, for abnormal glycosylation by IEF. In general, a fuller understanding of the disease processes may lead to insights into the mechanisms that control site occupancy in normal proteins and to the functions of oligosaccharides.

1.3 Carbohydrate Deficient Glycoprotein Syndrome

The carbohydrate-deficient disorders are a group of multisystemic diseases characterised by a deficiency of the oligosaccharides attached to glycoproteins which may result from one or more defects in glycosylating enzymes. CDGS belongs to this group of diseases, and was first reported by Jaeken et al in 1984. It is characterised by severe disorders of the nervous system, which include growth retardation and psychomotor retardation, abnormal distribution of adipose tissue, 'peau d'orange', squints and abnormal ocular movements and infertility. A large number of serum glycoproteins have been reported to be abnormal and these include transport proteins, glycoprotein hormones, complement factors, enzymes, and enzyme inhibitors. Amongst these the transport protein, transferrin, is an important diagnostic indicator of the disease. Transferrin is an iron transport protein containing bi- and tri- antennary complex glycans attached to Asn413LysSer and Asn611ValThr, both of which sites are normally fully occupied. The glycosylation of transferrin has been analysed by paper electrophoresis and P4 GPC and exoglycosidase sequencing of the released sugars (Yamashita et al 1989). It has also been analysed at the protein level by isoelectric focusing by Yamashita et al (1989) and within this present project by Dr. Brian Winchester. The IEF data indicate that normal human serum transferrin contains bi- and tri- antennary complex glycans +/- core fucose. However between 96-98% of the sites contain bi-antennary glycans with no core fucosylation. No neutral structures were detected, 4% of the sugars contained one sialic acid residue, 95% contained two and 1% contained three, and all the glycosylation sites were occupied. This indicates that normal transferrin may contain glycoforms with 2,3,4,5 and 6 sialic acid residues (designated S2-S6). In contrast, transferrin from CDGS patient typically contained up to 37% of glycoforms in which only one glycosylation site was occupied, and 13% in which neither site had been glycosylated (Yamashita et al 1992), although the sugars at the sites were indistinguishable in type and charge from the normal glycans. Therefore CDGS transferrin may contain glycoforms with 1,2,3,4,5 and 6 sialic acid residues (S1-S6). The set of fully sialylated glycoforms (S2-S6) is shown in figure 3 together with non-glycosylated transferrin.

In this study the glycoforms of normal and CDGS transferrin were resolved by CZE on the basis of charge. The structure of the glycoform populations was assigned by monitoring the elution positions of enzymatically modified transferrin, and a peak characteristic of CDGS was identified.

2. Materials and methods

2.1 High Performance Capillary Zone Electrophoresis (CZE) separations were carried out on a P/ACE System 2100 (Beckman Instruments) using a fused silica capillary (107cmx75μm ID). The detection was by UV at 200nm. Injections of 1-2mg/ml for 1.5-2 seconds were performed (5ng) under high pressure. The buffer was 50mM Tris, 50mM sodium glutamate (CH_3CH_2COONa) (pH 8.54-8.64). The voltage was 10kV.

2.2 Enzyme digestions

(i) Arthrobacter ureafaciens neuraminidase (Oxford GlycoSystems):

The time course was performed over 260 minutes using an optimised dilution of the *Arthrobacter ureafaciens* neuraminidase at 4mU:2mg transferrin/200μl 100mM sodium acetate pH5.0 at 37°C. Aliquots were removed from the reaction mixture during the digestion and the reaction stopped by freezing at -20°C. *Arthrobacter ureafaciens* neuraminidase cleaves NeuNAcα2-6>3,8 R

(ii) Streptococcus pneumoniae β-galactosidase (Oxford Glycobiology Institute): digestion of asialo glycans: 2mg/ml of substrate were incubated with 300mU/ml of β-galactosidase in 50mM citrate/phosphate buffer pH6 for 18h at 37°C. The enzyme cleaves Galβ1-3,4>6 GlcNAc/R

(iii) Streptococcus pneumoniae β-galactosidase + *Streptococcus* pneumoniae β-hexoseaminidase (Oxford Glycobiology Institute)*:* 2mg/ml of substrate were incubated with 300mU/ml of β-galactosidase + 500mU/ml of β-hexoseaminidase in 50mM citrate/phosphate buffer pH6. Incubation was for 18h at 37°C. *Streptococcus pneumoniae* β-hexoseaminidase cleaves GlcNAc β1-2,4,6 Man or Glcβ1-3,6Gal

(iv) Streptococcus pneumoniae β-galactosidase + *Streptococcus pneumoniae* β-hexoseaminidase + *Streptococcus (diplococcus) pneumoniae* Endoglycosidase D (Boehringer Co.)*:* 2mg/ml of substrate was incubated with 300mU/ml of β-galactosidase + 500mU/ml of β-hexoseaminidase + 150mU of endoglycosidase D for 18h at 37°C in 50mM citrate/phosphate buffer pH6. Endoglycosidase D cleaves between the GlcNAc residues in the chitobiose core.

2.3 Substantially iron free normal human transferrrin (T-0519) was obtained from Sigma Chemical Company.

3. Results

3.1 Resolution of normal serum transferrin glycoforms by CZE

Normal transferrin was separated into five glycoform populations (Fig.4) by CZE using Tris buffer and sodium glutamate (pH 8.54 - 8.64). Digestion of the normal population with *Arthrobacter ureafaciens* neuraminidase reduced all the glycoforms to one major and two minor asialo populations (Fig.5: 260 mins). Consistent with the findings of Yamashita et al (1990) these data confirm that normal serum transferrin contains no neutral glycoforms. The data indicate that transferrin consists of a population of glycoforms which have been resolved according to the number of sialic acid residues they contain. Approximately 95-98% of the glycoforms of normal human transferrin contain two di-sialylated bi-antennary complex sugars and a charge due to four sialic acid residues (S4) (Yamashita et al 1990). Analysis of commercial preparations, such as was used here in order to detect a range of glycoforms, have shown that the percentages of the glycoforms supplied are variable and depend on the manufacturing process, but the major glycoform is invariably S4 (Yamashita et al 1989; Van Pelt et al 1987; Spik et al 1975). On the basis of these previous analyses therefore, Peak 4 was assigned to S4, and the remaining peaks were assigned to transferrin carrying bi- or tri-antennary complex glycans with 2,3,5 or 6 sialic acid residues. The percentages of each peak (S2-S6) were 2%:12%:67%:13%:6% (Table I).

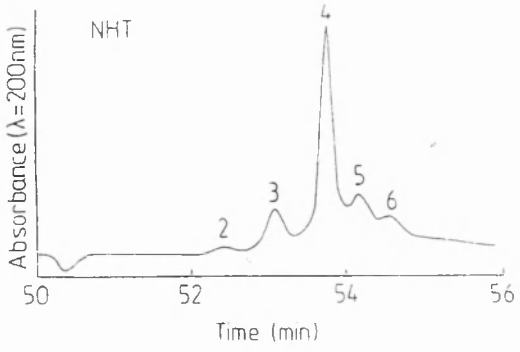


Figure 4: Capillary electrophoresis separation of glycoforms of normal human transferrin. Peak 4 (S4) contains the major glycoform which is glycosylated with two di-sialylated biantennary complex glycans, the other peaks represent transferrin with 2,3,5 or 6 sialic acid residues attached to bi- or tri- antennary complex glycans.

3.2 Time course showing the successive removal of sialic acid residues from normal human serum transferrin

The CE apparatus was arranged with the anode at the injection end of the capillary, therefore neutral glycoforms eluted first and the more negatively charged populations eluted later. These data show that all of the partial and final digestion products eluted earlier than any of the original glycoforms, indicating that all the populations have become less negatively charged due to the loss of sialic acid. The time course (Fig. 5) shows that digestion with *Arthrobacter*

ureafaciens neuraminidase has successively removed sialic acid residues from the native population of S2-S6 glycoforms. The original five sialylated glycoforms have been modified to one population containing only neutral glycans.

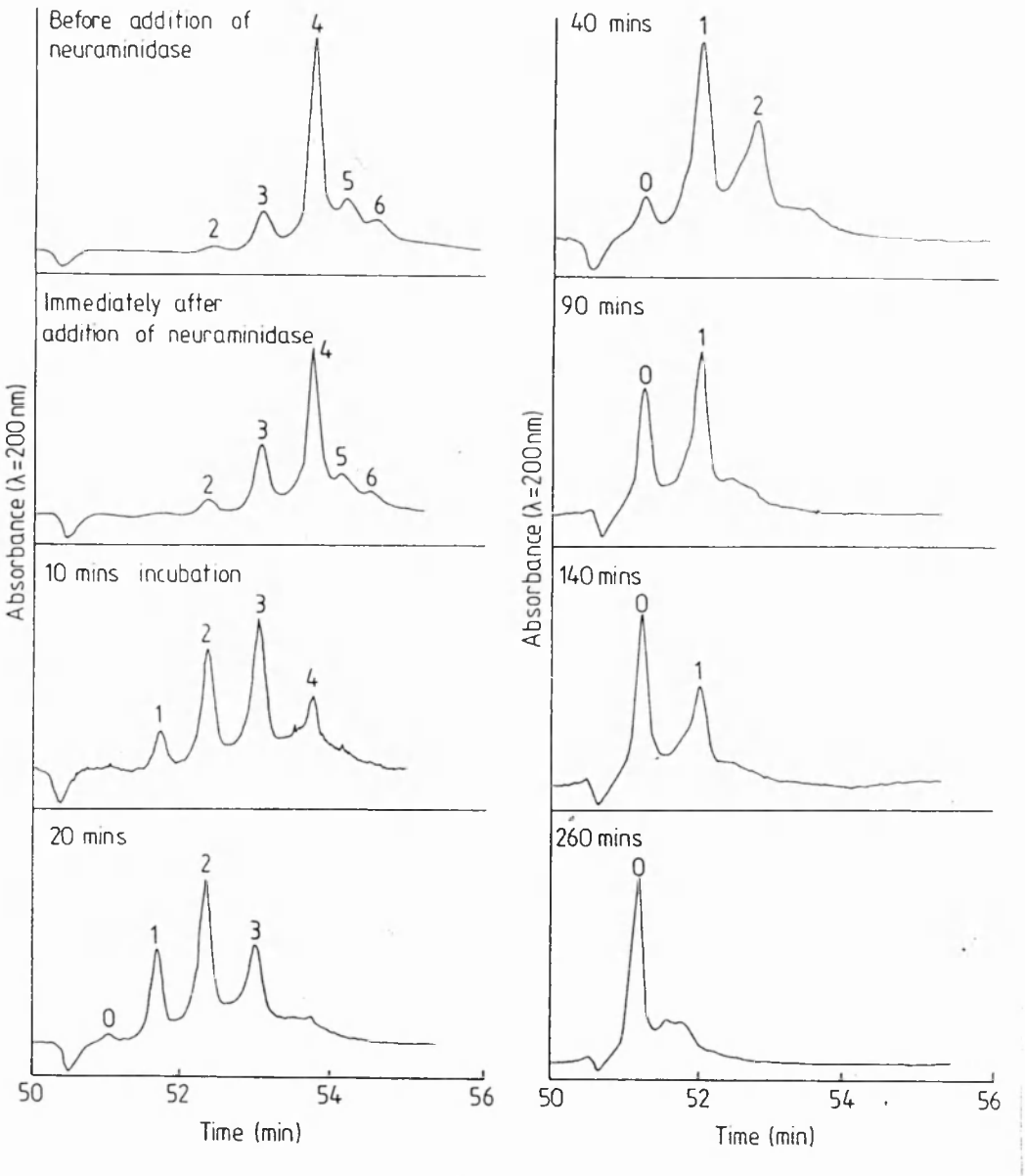


Figure 5: The neuraminidase digestion of the glycoform populations of normal human transferrin monitored by CE.

3.3 CZE analysis of sequential exoglycosidase digestions of the asialo glycoforms of human transferrin

The glycan structures attached to asialo transferrin were sequentially removed in a series of mixed exoglycosidase digestions shown schematically in figure 7. Following the neuraminidase digestion, incubation with β-galactosidase removed terminal galactose residues from the

desialylated population, next β-hexoseaminidase removed terminal N-acetyl glucosamine, and finally endo D cleaved the remaining sugar within the chitobiose core leaving a single GlcNAc residue attached to the protein. The final sugar residue was not removed because cleaving the asparagine sugar linkage with PNGase F would convert the asparagine residue to aspartic acid and introduce a charge to the glycoform, altering the basis of this particular CE separation. The larger sugars elute earlier because they contain more hydroxyl groups and can therefore acquire more negative charge by complex formation with the glutamate ions in the buffer.

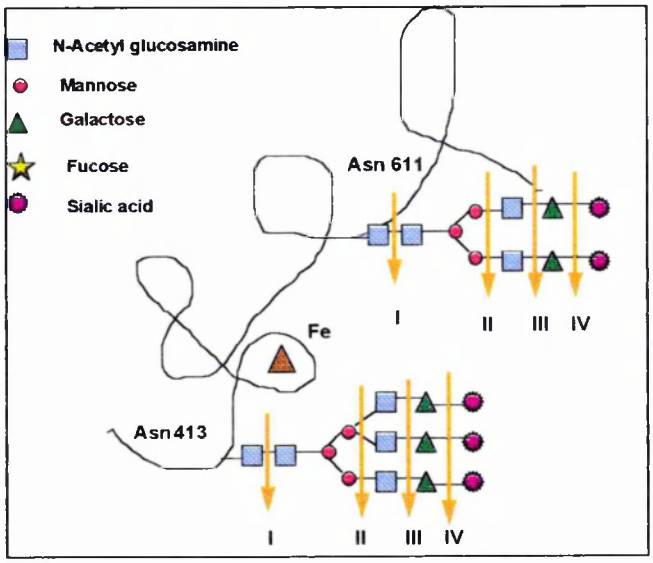


Figure 6: The C_t domain of one of the glycoforms of transferrin (S5) *showing the cleavage positions of the exoglycosidase enzymes used in the sequential digestion. The enzymes were (I) Arthrobacter ureafaciens neuraminidase (II) Jack Bean β-galactosidase (III) Jack Bean β-hexoseaminidase (IV) Endoglycosidase D. Transferrin consists of two domains, the N_t domain is not glycosylated.*

This analysis (Fig. 7) was consistent with previous studies (Yamashita et al 1989, 1990) which show that the major population of transferrin glycoforms contains bi-antennary complex glycans. After each enzyme digestion the major glycoform eluted later, indicating that the glycans had been sequentially reduced in size by the specific enzyme in each of the digests. The analysis therefore defined the elution positions of the series of transferrin glycoforms (I-IV) modified by the glycosidase enzymes as shown schematically in Fig. 6. The data indicate that, within one charge band, this CZE method can resolve transferrin glycoforms on the basis of size. In addition, the oligosaccharides were able to be sequenced while attached to the protein and without denaturing (which would cause a loss of resolution). This indicates that the oligosaccharides attached to transferrin are accessible to the glycosidase enzymes. Importantly, this analysis established the elution position of a new transferrin glycoform containing only one GlcNAc residue at each site (IV).

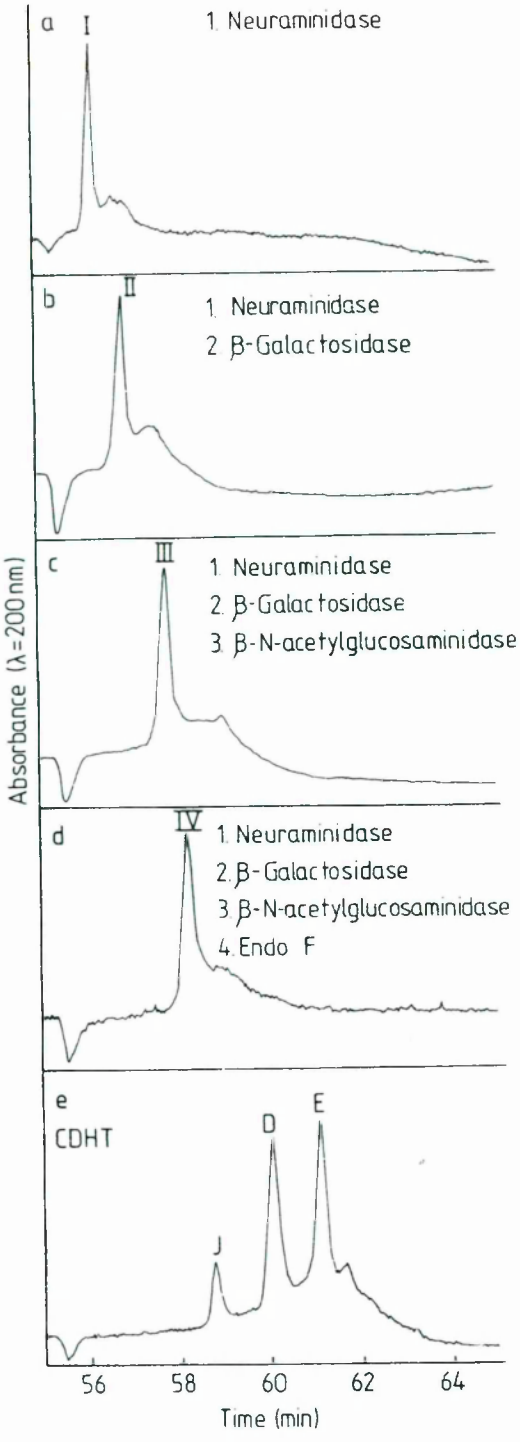


Figure 7: (a-d) The sequential digestion of the oligosaccharides attached to normal human transferrin showing the elution positions of the series of glycoforms generated by the digestions. (e) A sample of CDGS transferrin included for comparison (see 3.4)

3.4 CE analysis of transferrin isolated from two patients with CDGS.

Figure 8 shows the analysis of transferrin from two patients (b,c) compared with normal transferrin (a). A co-injection of the CDGS sample in (b) with normal transferrin analysed in (a) is shown in figure 5d. These data indicate that both the CDGS samples contain S4, S5 and S6 in approximately the same proportion as transferrin from normal serum (see table 1). However they also contain increased levels of S2, and a population, J, which does not align with any of

the populations previously identified in this study. The elution profile of this sample has also been aligned with the data in figure 7. This indicates that J elutes later, and is therefore smaller, than any glycoform in the exoglycosidase digestion series, including the population containing only a single N-acetyl glucosamine residue at each site. J also elutes earlier than any of the charged glycoforms (Fig. 5), indicating that it does not contain sialic acid. This suggests that J may contain non-glycosylated transferrin, the presence of which is a characteristic of CDGS transferrin.

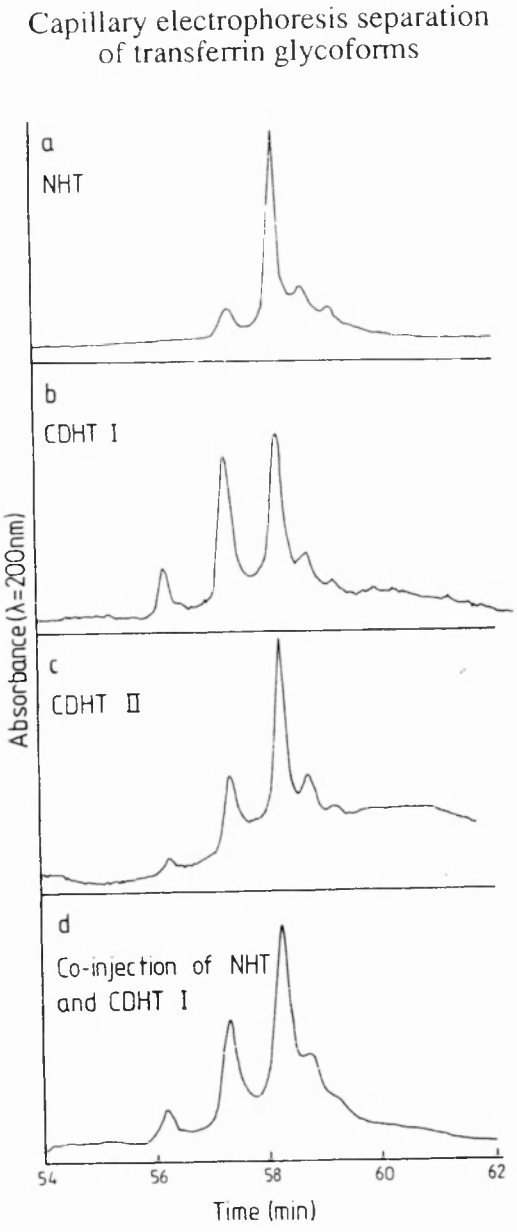


Figure 8a-d: The analysis of transferrin from two patients with CDGS (b,c) compared with normal transferrin (a). A co-injection of the CDGS sample in (b) with normal transferrin from (a) is shown in figure 8d.

Percent ratios of transferrin glycoforms						
	S0	**S2**	**S3**	**S4**	**S5**	**S6**
NHT (Sigma)	0	2	12	67	13	6
CDHT I	11	34	4	41	8	2
CDHT II	4	22	0	57	14	5

Table 1: The ratio of transferrin glycoforms in normal and CDGS transferrin

3.5 CE analysis of desialylated CDGS

Figure 9 shows an overlay of the separations of the natural and asialo populations of CDGS transferrin. Peak E contains the fully sialylated bi-antennary complex glycoform (S4). The data indicate that B contains transferrin with only one glycosylation site occupied and is the disialylated form, S2. E elutes later than B, consistent with the finding that glycoforms which contain more sialic acid elute later in the separation (*3.2).* E' is the asialo form of E and B' is the asialo form of B; E' is larger than B' and elutes earlier, consistent with the trend established earlier which demonstrated that the larger neutral glycoforms of transferrin elute earlier in the separation (*3.4).* These data support the assignment of B to a glycoform with only one occupied site containing a fully sialylated bi-antennary complex glycan (S2). However it should be noted that the data do not exclude the possibility that B may contain some transferrin with two monosialylated biantennary complex sugars, on desialylation this glycoform would elute with E'.

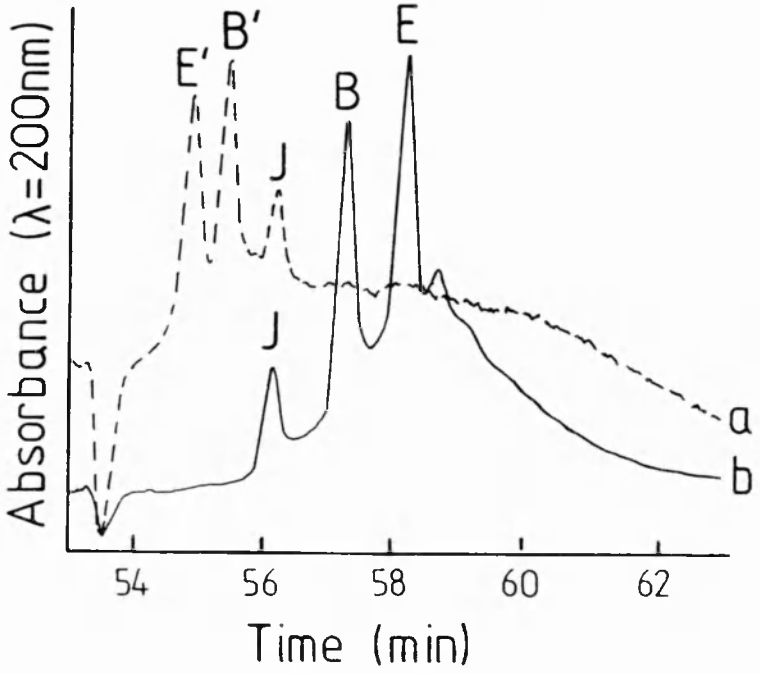


Figure 9: An overlay of CDGS transferrin and the same sample after de-sialylation with *Arthrobacter ureafaciens* neuraminidase.

4. Discussion:

4.1 The glycoforms of transferrin have been resolved on the basis of charge by CZE

In this study, using Tris/glutamate buffer at pH 8.6, CZE has been used to the resolve the glycoforms of transferrin quantitatively at the protein level on the basis of charge. The charge was shown to be due to terminal sialic acid residues which were removed by digestion with *Arthrobacter ureafaciens* neuraminidase reducing the whole population to a single glycoform. Interestingly, the analysis of erythropoetin by CZE (Tran et al) also showed that neuraminidase could reduce the multiple peak structure to a single peak. This suggests that CE resolution of glycoforms based on the differences in sialylation may be a useful approach to the analysis of other sialylated glycoproteins. In this study, in figure 4, the glycoform populations S2-S6 were assigned to molecules carrying from 2-6 sialic acid residues on the two glycans attached to each molecule of transferrin, by analogy with previous studies of transferrin isolated from human serum (Yamashita et al 1989). The relative proportions of the sialylated populations were S2:2%, S3:12%, S4:67%, S5:13%, S6:6%. The neuraminidase digestion was followed by CZE, and this allowed the variations in the relative proportions of the different glycoforms to be monitored throughout the digestion. In a similar study involving the digestion of the oligomannose glycans attached to RNase B, the Man6 glycoform population was found to be more resistant to the mannosidase enzyme than the others (Rudd et al 1992), suggesting that the protein gives some protection to this particular glycoform in which the terminal mannose residue is closer to the chitobiose core. In the case of transferrin, all of the sialic acid residues are terminal and equidistant from the chitobiose core. Although the residues were removed sequentially and all glycoform populations from S6 to S0 were observed during the course of the digestion, there was no indication that any particular sialylated glycoform was more resistant than the others, suggesting that all the sialic acid residues are equally accessible to *Arthrobacter ureafaciens* neuraminidase. Interestingly, although the assay system clearly established elution positions for variably sialylated glycoforms, it did not resolve the peaks further into glycoforms carrying the same charge but with a different mass. For example, S3 which is derived from the loss of one SA residue from transferrin with two bi-antennary complex glycans, is not distinguished from S3 which is derived from the loss of two SA residues from a transferrin molecule carrying one tri- and one bi-antennary complex glycan.

4.2 Sequential exoglycosidase digestions of glycans attached to intact undenatured transferrin

The results of sequential exoglycosidase digestions (Fig.7) of the intact, undenatured glycoprotein were consistent with previous studies (Yamashita et al 1989, Spik et al 1975, and Van Pelt et al 1987) which indicated that the major population of sugars attached to normal human transferrin was the sialylated bi-antennary complex glycan, SA2H5N4 (glycan composition formula of oligosaccharide containing 2xsialic acid, 5xhexose and 4xN-acetyl hexosamine glycans). Sequential digestions removed sialic acid, galactose and N-acetyl glucosamine in turn and the modified glycoforms took longer to elute as their size decreased.

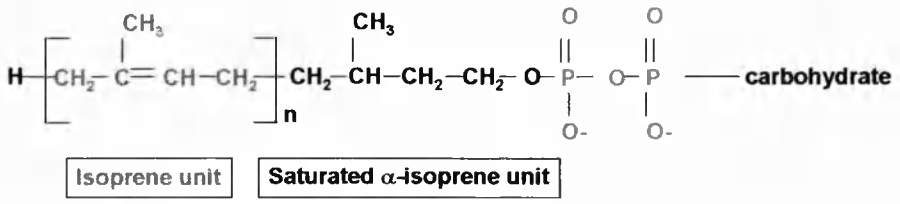
These data indicate that, within the neutral charge band, the system is capable of resolving glycoforms on the basis of size.

4.3 Glycosylation profile of CDGS transferrin

Separation of CDGS transferrin by CZE (Fig. 7e and 8) indicates that both patients have increased levels of S2 and also a population of glycoforms, J, not present in either in normal or desialylated normal transferrin. J elutes earlier than the S0 glycoforms and later than transferrin containing only two single GlcNAC residues attached to the protein. The elution position suggests that it may contain non-glycosylated transferrin. This would be consistent with previous studies of CDGS transferrin, which has been shown to contain non-glycosylated and monoglycosylated molecules. The increase in S2 may be due to an increase in monosialylated transferrin, but this was not found by Yamashita, and suggests that the increase is more probably due to the presence of transferrin glycoforms in which only one glycosylation site has been filled with a fully sialylated bi-antennary complex sugar. This would also be consistent with previous studies of CDGS transferrin. The data suggest that profiling transferrin under the conditions described in this chapter would reveal increased levels of S2 and S0, indicative of CDGS.

4.4 *Variable sequon occupancy: factors which control protein glycosylation*

Glycoforms which contain unglycosylated sites may arise for several reasons. For example in the previous chapter the possibility that ribonuclease A and B may arise from a competition between glycosylation and folding was discussed. In addition, natural sialidases have been identified in mammalian systems (Suzuchi et al 1993) which may modify the sugar side chains of glycoproteins by cleaving the asparagine-sugar linkage. In the case of CDGS transferrin the data do not support the latter possibility. Following enzyme cleavage by PNGase, asparagine residues are converted to aspartic acid and the introduction of two additional charges to the transferrin molecule would alter the CE elution position, so that the glycoform would not be expected to elute with neutral structures. In addition, studies of one peptide N-glycosidase (PNGase F) in this thesis (ch.3,4) have suggested that it is specific and does not cleave N-linked glycans from all native glycoproteins. In CDGS every glycoprotein which has been analysed so far has been deficient in glycosylation. It is therefore possible that the sites in CDGS transferrin were not glycosylated in the endoplasmic reticulum (ER). Some general rules have emerged with respect to the factors which control the attachment of oligosaccharides to potential glycosylation sites in the ER. For example, to be glycosylated an asparagine residue must form part of the tripeptide sequon AsnXSer/Thr where X is any amino acid apart from proline, although the presence of this sequon is not in itself sufficient to ensure glycosylation. Although the same glycosylation machinery is available to all the proteins which are translated in a particular cell and use the secretory pathway, it has been estimated that between 10% and 30% of potential glycosylation sites are not occupied (Gavel and von Heijne 1990). The three dimensional structure of the individual protein clearly has a role in determining the extent of its



Dolichol phosphate

Figure 6: The lipid component of the dolichol phosphate precursor oligosaccharide

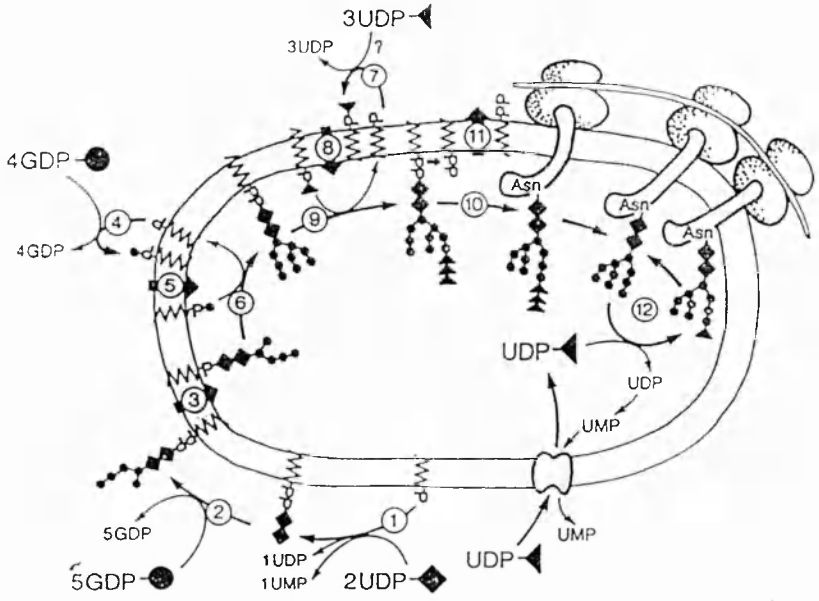


Figure 11: The synthesis of the dolichol phosphate precursor takes place first on the cytoplasmic face and secondly on the lumen face of the ER.

glycosylation (Goochee et al 1992). A number of mechanisms may be involved, including the position of the glycosylation site in the protein. However, in the case of CDGS none of these factors would seem to explain the widespead lack of glycosylation.

Another factor which controls the proportion of occupied sites is the availability of the dolichol pyrophosphate linked $Man_9GlcNAc_2Glc_3$ precursor oligosaccharide. The lipid component of this precursor is dolichol, a long chain isoprenol of 14-20 isoprene units (Fig.10) which anchors the growing oligosaccharide in the lipid bi-layer of the ER. Initially the sugar chain is elongated in the cytoplasm until a Man 5 glycan is attached at which point the molecule 'flips' so that the remaining processing of the Man 5 glycan to the $Man_9GlcNAc_2Glc_3$ precursor takes place within the lumen of the ER (Fig.11). The precursor is transferred to the asparagine residue in the glycosylation sequon of the nascent glycoprotein by an oligosaccharyl transferase which is situated 30-40A above the membrane of the ER and oriented roughly parallel to the membrane surface (Nilsson et al 1993). There are many stages in this process which may lead to abnormal glycoprotein metabolism and result in the incomplete glycosylation of a wide range of glycoproteins (Jaeken et al 1993).

4.5 Alteration of glycosylation patterns with disease

Alteration of the glycosylation patterns of natural glycoproteins have been noted in some diseases. In this thesis, serum IgG isolated from a rheumatoid arthritis (RA) patient was found to contain increased levels of agalactosyl IgG (Chapter 6), consistent with the findings of others (Parekh et al 1985) which established that in RA the distribution of the glycoform populations in the Fc is altered, reflecting an increase in the proportion of N-linked glycans terminating in N-acetyl glucosamine. Interestingly, incomplete glycosylation has also been noted in serum transferrin as a result of alcohol abuse (Stibler et al 1986), in retinitis pigmentosa (Andreasson et al 1992) and in hepatocellular carcinoma (Yamashita et al 1989).

4.6 Applications of CE in glycobiology:

Once a method has been developed for fully or even partially resolving a glycoform mixture by capillary electrophoresis many investigations become possible. Some of these applications have been explored in this thesis in this chapter or in the study of RNase B (ch.4). These include: (i) resolving glycoforms, without destruction, at the protein level and enabling the preparation of individual glycoforms (ii) following details of enzymatic modifications to glycosylation directly during the course of the digestions (iii) probing alterations to glycosylation associated with disease (vi) following the enzymatic release of glycans from the protein (vii) exoglycosidase sequencing of oligosaccharides while attached to the protein (vii) comparing the stability of different glycoforms to protease activity (vii) examining the role of the protein in protecting the sugars from enzymatic digestion by comparing the susceptibility of the oligosaccharides bound to the protein to the same sugars in free solution. In addition the profiling of tPA (Taverna et al 1992) suggests that the technique may be useful for monitoring

alterations to the glycoform populations during manufacturing processes and comparing the glycosylation of glycoproteins from different cell lines.

5. References to chapter 5:

Andreasson, S., Stibler, H. and Ehinger, B. (1992) Acta Opthalmologica 525-527 Serum transferrin in patients with retinitis pigmentosa.

Gavel, Y. and von Heijne, G. (1990) Protein Engineering 3, 433-442 Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering.

Goochee, C.F., Gramer, M.J., Anderson, D.C., Bahr, J.B., and Rasmussen, J.R. (1992) Frontiers in Bioprocessing II, pp199-240 Eds Todd, P., Sikdar, S.K. and Bier, M. Pubs. A.C.S.

Jaeken, J., van Eijk,H.G., van der Heul, C., Corbeel, L., Eeckels, R. and Eggermont E (1984) Clin. Chim. Acta. 144 245-247 Sialic acid deficient serum and cerebrospinal fluid transferrin in a newly recognised genetic syndrome.

Jaeken, J., Carchon, H., and Stibler, H. (1993) Glycobiology 3 423-428 The carbohydrate-deficient glycoprotein syndromes: pre-Golgi and Golgi disorders.

Kornfeld, R. and Kornfeld, S. (1985) Ann. Rev. Biochem 54 631-634 Assembly of asparagine linked oligosaccharides..

Nilsson, I.M. and von Heijne, G. (1993) J.Biol. Chem. 268 5798-5801 Determinations of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane.

Parekh, R.B., Dwek, R.A., Sutton, B.J., Fernandes, D.L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, K., Takeuchi, f., Nagano, Y., Miyamoto, T. and Kobata, A. (1985) Nature 316, 452-457 Association of Rheumatoid Arthritis and Primary Osteoarthritis with changes in the glycosylation pattern of total serum IgG.

Rudd, P.M., Scragg, I.G., Coghill, E. & Dwek, R.A. (1992) Glycoconjugate Journal 9, 86-91 Separation and analysis of the glycoform populations of RNase B using capillary electrophoresis.

Spik, G., Bayard, B., Fournet, B., Strecker, G. Bouqualt, S. and Montreuil, J. (1975) FEBS Lett. 50 296-299 Studies on glycoconjugates LXIV Complete structure of two carbohydrate units of human serotransferrin.

Stibler, H., Borg, S. and Joustra, M. (1986) Alcoholism Clinical and Experimental Research 10 535-544 Micro Anion Exchange Chromatography of carbohydrate deficient transferrin in serum in relation to alcohol consumption.

Suzuchi, T., Seko, A., Kitajima, K., Innoue, Y. and Innoue, S. (1993) Biochem. and Biophys. Res. Comms. 194 1124-1130 Identification of peptide:N-glycanase activity in mammalian-derived cultured cells.

Taverna, M., Baillet, A., Biou, D., Schlutter, M., Werner, R., Ferrier, D. (1992) *Electrophoresis* 1992 13, 359-366

Tran, A.D., Huynh, O.T., Park, S., Ryall, R.R., Lisi, P.J. and Lane, P.A. 1991 *J. Chromatog.* 524, 459-471 Separation of carbohydrate-mediated microheterogeneity of recombinant human erythropoietin by free solution capillary electrophoresis: effects of pH and buffer type.

Van Pelt, J., Damm, J., Kamerling, J.P. and Vliegenthart, J.F.G. (1987) Carbohydr. Res. 169 43-51 Separation of sialyl-oligosaccharides by medium pressure anion exchange chromatography on mono-Q.

Yamashita, K., Koide, N., Endo, T., Iwaki, Yo and Kobata, A. (1989) J. Biol. Chem. 264 2415-2423 Altered glycosylation of serum transferrin of patients with hepatocellular carcinoma.

Yamashita, K., Ideo, H., Ohkura, T., Fukushima, K., Yusa, I., Ohno, K. and Takeshita, K. (1993) J. Biol. Chem. 268 5783-5789 Sugar chains of serum transferrin from patients with carbohydrate deficient glycoprotein syndrome.

Chapter 6

Novel strategies to determine the glycosylation of IgG and to prepare pure agalactosyl glycoforms of IgG, Fab and Fc.

Publications associated with this chapter:

(1) Harvey, D.J. and Rudd, P.M.; Bateman, R.H., Bordoli, R.S., Howes, K., Hoyes, J.B. and Vickers, R.G. (1994) Organic Mass Spectrometry 29 753-765 Examination of Complex Oligosaccharides by Matrix-Assisted Laser Desorption Mass Spectrometry on Time-of-Flight and Magnetic Sector Instruments

(2) Malhotra, R., Wormald, M., Rudd, P.M., Fischer, P.B., Dwek, R.A. and Sim, R.B. (1995) Nature Medicine 1 237-241 Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose binding protein

Acknowledgements:

1. The preparation and analysis of IgGRF was carried out by PMR in collaboration with Dr. Ian Scragg before the registration of this thesis. The sucrose density gradient conditions (SDG), RF assay protocol (modification of method by Hay et al (1975)) and RF preparative conditions (modification of the method of Nardella et al (1981)), were initially developed by PMR. The monomeric IgG, the chemically cross linked IgG oligomers and the IgA standard were prepared by PMR. The particular RF preparation, RF assay and SDS PAGE analysis of RF aggregates used in section 4.3.2-3 was carried out by Dr. I. Scragg.
2. The mass spectrometry studies were carried out in conjunction with Dr. David Harvey
3. HPLC technology was developed by Mr. Geoffrey Guile
4. The molecular modelling was carried out in conjunction with Dr. Mark Wormald
5. The mannose binding and C4 fixation plate assays were performed by Dr. R. Malhotra

Abbreviations: IgG: Immunoglobulin G; H: heavy chain; L: light chain; Fab: the antigen binding fragment of IgG; Fc: crystallisable fragment of IgG; RF: rheumatoid factor; MBP: mannose binding protein; CRD: carbohydrate recognition domain; G2,G1,G0: complex biantennary oligosaccharides containing 2,1 or 0 arms respectively terminating in galactose. The arms which lack galactose terminate in N-acetyl glycosamine; P4 GPC: P4 Gel Permeation Chromatography; WAX: weak anion exchange chromatography; CAT: cation exchange chromatography; MALDI MS: Matrix Assisted Laser Desorption Mass Spectrometry; MASP: mannose binding protein - associated serine protease; B- bisecting N-acetyl glucosamine; F: fucose; H: hexose: N: N-acetyl hexosamine

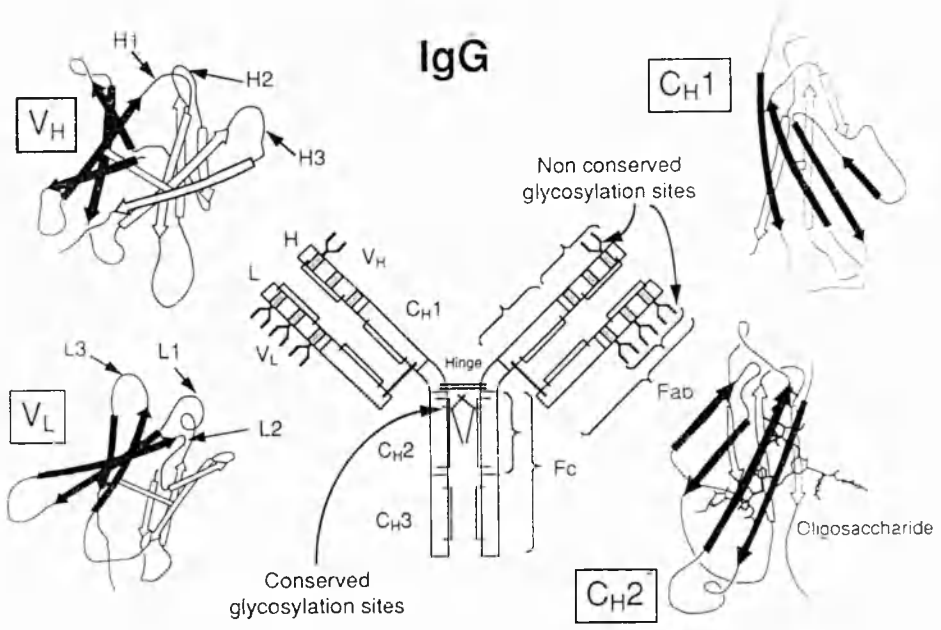
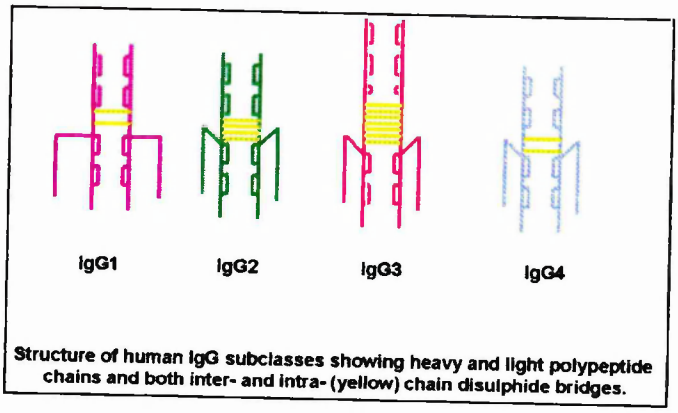


Figure 1: The antibody molecule *consists of two heavy (H) and two light (L) chains, linked by disulphide bridges (solid lines) and divided into homologous regions of sequence (V_H, C_H1, C_H2, C_H3, V_L and C_L) each of which has an intrachain disulphide bridge. The folding of the chain within each domain is shown schematically, C_L and C_H3 fold in a very similar way to C_H1. Arrows represent strands of β-sheet, and the solid bars are the intrachain disulphide bridges which link the two β-sheets of each domain. The homologous four stranded 'faces' of each domain are shown cross hatched. This face in C_H2 is partially covered by the branched oligosaccharide chain, but it forms the domain pair interface in C_H1:C_L and C_H3:C_H3. In V_H and V_L the stippled segments represent the hypervariable regions of sequence (H1, L1 etc) which, in the 3-dimensional structure, together form the antigen binding site (Dwek et al 1984).*

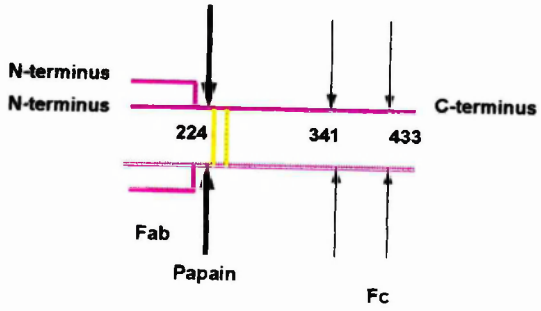
Novel strategies to determine the glycosylation of IgG and to prepare pure agalactosyl glycoforms of IgG, Fab and Fc.

Abstract

Immunoglobulin G (IgG) consists of two heavy (H) and two light (L) polypeptide chains. It can be cleaved by papain into two fragments of Fab (the antigen binding fragment) and one of Fc (the region which mediates many of the effector functions). IgG contains two conserved N-linked glycosylation sites at Asn 297 on each of the CH2 domains which are located in the Fc. In addition 20% of IgG molecules are also glycosylated in the hypervariable region of the Fab. In this chapter a novel strategy was developed to determine the glycosylation of IgG involving anion exchange chromatography with on-line fluorescent detection and simultaneous sequencing of multiple oligosaccharides analysed by matrix assisted laser desorption ionisation mass spectrometry. The sugars were released from human serum IgG by hydrazinolysis and desialylated with *Arthrobacter ureafaciens* sialidase. The pool of asialo oligosaccharides was analysed without separation of the individual components using a novel combination of exoglycosidase digestions, ion exchange chromatography and mass spectrometry. Consistent with previous analyses of human serum IgG a range of complex biantennary oligosaccharides were identified, confirming that IgG consists of a mixture of glycoforms. The distribution of charge on the released IgG glycans was determined using an optimised weak anion exchange chromatographic method. The same method was used to analyse the charged glycans released from serum IgG from a patient with rheumatoid arthritis (RA). The most notable difference in the charge profiles of the two IgG samples was that RA IgG contained two populations of monosialylated oligosaccharides (G1, G2) while normal IgG contained only one (G1). The released IgG glycans were digested with *Arthrobacter ureafaciens* sialidase, *Choronia lampas* fucosidase and Jack bean N-acetylhexosaminidase to reduce all the oligosaccharides to three populations differing only in the extent of their outer arm galactosylation. A comparison of the galactose content of normal and RA IgG, using cation exchange chromatography, indicated a decrease (13%) in G2 sugars and an increase (16%) in G0 sugars in RA IgG compared with normal IgG, while the G1 population remained relatively stable (36% in RA compared with 39% in normal). These data are consistent with the trends established in previous studies of RA IgG in this laboratory. In addition to the increased levels of IgG0, another characteristic feature of RA is the presence in serum and synovial fluid of IgG rheumatoid factors (RFs) which self-associate to form aggregates of IgG which may be either soluble or deposited in joints. RFs were shown to consist of IgG molecules which are glycosylated in the Fab region as well as in the Fc and contain mainly G0 type sugars. IgGRF, containing approximately 85% of IgG0 sugars, was used to test the possibility that the terminal N-acetyl glucosamine residues present on IgG0 glycoforms may be exposed and available for binding. One candidate for binding is the calcium dependent serum lectin mannose binding protein which binds pathogens with a high



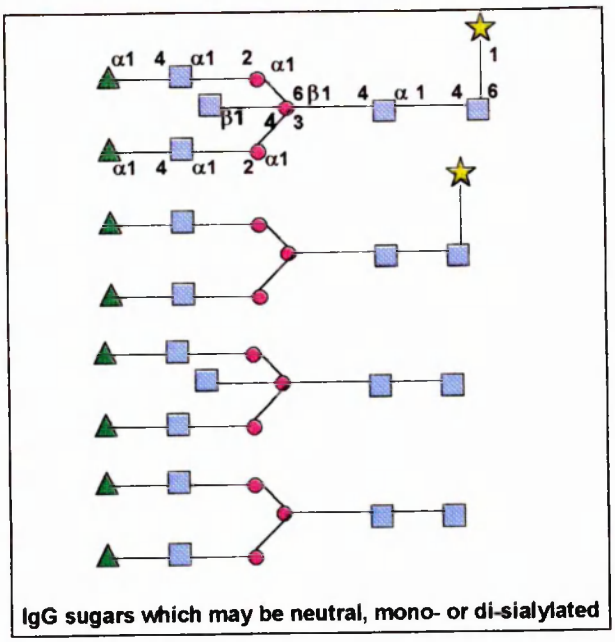
Structure of human IgG subclasses showing heavy and light polypeptide chains and both inter- and intra- (yellow) chain disulphide bridges.

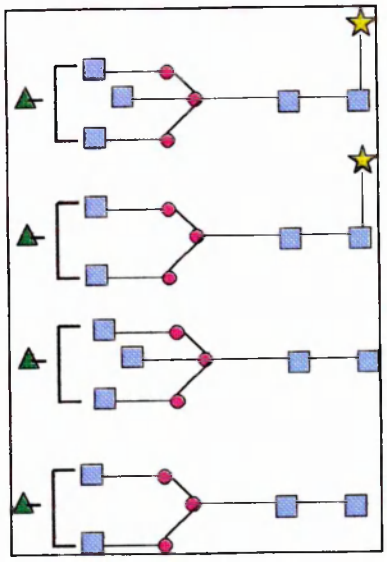


Papain cleavage points of human IgG1. The major cleavage site is between residues 224 and 225. Two minor cleavage sites are between residues 341/2 and 433/4

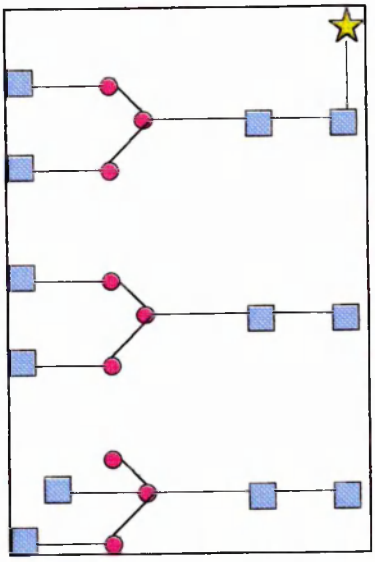
Figure 2a,b

concentration of terminal N-acetyl glucosamine residues on their surface. Preliminary studies explored the possibility that this lectin might also recognise terminal N-acetyl glucosamine residues on IgG0 when multiply presented. IgGRF (G0 approximately 85%) bound more strongly to the lectin than normal IgG (G0: 20%) when both were immobilised and concentrated on dot blots. To examine the preliminary finding more rigorously pure sets of IgG0, FabG0 and FcG0 glycoforms were prepared. Normal human serum IgG was cleaved by papain into Fab and Fc, and the natural mixtures of IgG, Fab and Fc glycoforms (containing G2, G1 and G0 glycoforms) were modified to populations containing 100% G0 type sugars by removing sialic acid and galactose residues with *Arthrobacter ureafaciens* and *Streptococcus 6646K* β-galactosidase. Subsequent analysis showed that IgG0 and FcG0 but not FabG0 bound mannose binding protein when multiply presented on a microtitre plate.





IgG sugars which may be neutral or monosialylated



IgG neutral glycans

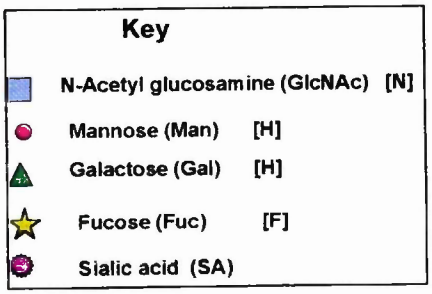


Figure 3: Primary oligosaccharide sequences of the major sets of N-linked sugars associated with IgG

Novel strategies to determine the glycosylation of IgG and to prepare pure agalactosyl glycoforms of IgG, Fab and Fc.

1. Background:

1.1 The structure of Immunoglobulin G

IgG (M.Wt. 150-160kD) consists of two heavy (H) and two light (L) chains linked together by disulphide bridges and divided into homologous regions of sequence (V_H, C_H1, C_H2, C_H3, V_L and C_L) (Fig. 1). V_H and V_L contain hypervariable regions of sequence which together form the antigen binding site. X-ray crystallography has shown that each region of homology in the IgG molecule corresponds to a compact, independently folded unit, and that these are linked together by short sections of extended polypeptide chain. Each immunoglobulin domain consists of two β-pleated sheets with antiparallel strands connected by loop regions. This general pattern is known as the immunoglobulin fold (Amzel and Poljak 1979) (ch.2). In humans there are four subclasses, distinguishable from each other by the arrangement of disulphide bridges in the hinge region of the molecule (Fig. 2a). The relative proportions of the different subclasses in polyclonal IgG are approximately IgG1: 70%, IgG2: 20%, IgG3: 7%, IgG4: 3%. IgG can be cleaved by papain into Fab (the antigen binding fragment) and Fc (the region responsible for the effector functions of the molecule) (Fig. 2b). The major cleavage site, on the C-terminal side of Lys 224, (Fig. 2b) is in the variable hinge region of the molecule and different subclasses of IgG show marked differences in susceptibility to the enzyme. IgG3 is the most susceptible followed by IgG1>IgG4>IgG2.

In IgG the two CH2 domains each contain a conserved glycosylation site at Asn297 to which complex biantennary oligosaccharides are covalently linked. On average 2.5+/-0.3 N-linked sugars are attached to each IgG molecule (Chang, S.C. 1993) of which two are located in the Fc. The additional oligosaccharides are found in the hypervariable regions in the Fab. IgG is associated with at least 30 different complex glycans (Fig. 3). The glycosylation pattern of IgG is species specific and may be influenced by physiological state (Rook et al 1991) or by cell culture conditions (Goochee et al 1992). Structural differences between the glycans arise primarily from sialylation, core substitutions and outer arm galactosylation.

(i) Sialylation

In normal human serum IgG typically 81% of the structures are neutral, 16% contain one sialic acid residue and 3% are disialylated (Rademacher et al 1985). All the sialic acid residues are linked α–2,6 to galactose and no other charged species have been detected. Most of the disialylated oligosaccharides are located in the Fab region (Rademacher et al 1985). Restricted access of the α2,6 sialyl transferase to the sugars within the CH2 domains may explain the low percentage of disialylated structures in the human Fc.

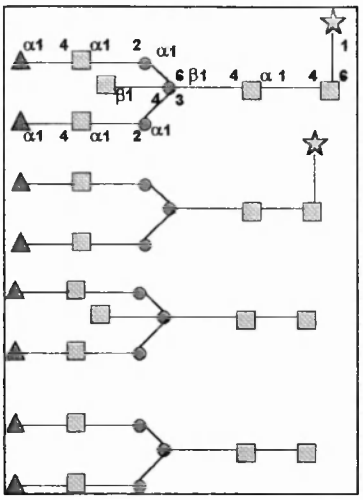


Figure 4: The four core structures (+B+F, -B+F, +B-F, -B-F) of the N-linked oligosaccharides in IgG

B: bisecting N-acetylglucosamine, F: fucose

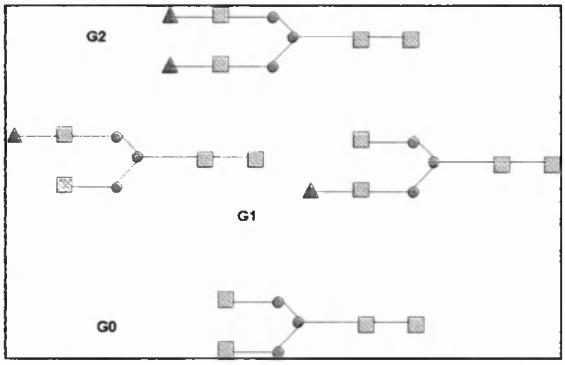


Figure 5: The G2, G1 and G0 classes of sugars released from IgG *(sialic acid, core fucose and bisecting GlcNAc have been removed by enzymatic digestion)*

Structural analysis of the glycans released by hydrazine from normal human serum IgG shows that all are of the complex bi-antennary type. They may be classified in several ways:

(i) according to the substitutions in the trimannosyl core. There are four different core structures (Fig. 3a) arising from the presence or absence of α1,6 fucose attached to the primary GlcNAc and β1,4-(bisecting) N-acetyl glucosamine attached to the second GlcNAc residue.

(ii) according to the number of terminal sialic acid residues. IgG sugars may be either neutral, mono- or di-sialylated (Fig. 4)

(iii) according to the number of galactose residue in the outer arm (Fig. 5)

(ii) Core substitution of fucose and N-acetylglucosamine

The addition of core fucose (F) and/or bisecting N-acetyl glucosamine (B) to the primary and secondary GlcNAc residues in the trimannosyl core allows four alternative core structures, +F+B, +F-B, -F+B, and -F-B (Fig. 4).

(iii) Outer arm N-acetyl glucosamine

Depending on the extent of galactosylation the complex glycans contain 0, 1 or 2 terminal N-acetyl glucosamine residues. Interestingly, polyclonal IgG has been found to contain a low percentage (3%) of structures with only one GlcNAc substitution on the trimannosyl core (Jefferis et al 1990). Invariably this was linked to the α1,3 arm, suggesting that the levels of GlcNAc transferase II (which adds GlcNAc α1,2 to the mannose on the α1,6 arm) may not always be sufficient to add GlcNAc to the 1,6 arm of every developing glycan.

(iv) *Outer arm galactosylation*

The outer arms of the biantennary sugars are classified into three types: those which contain a galactose residue on each arm (G2), those which contain one, either on the α1,3 or α1,6 arm (G1(α1,3) and G1(α1,6)) and those which contain no galactose (G0) (Fig. 5). NMR relaxation studies suggest that, for the glycans lacking galactose on the α1,6 arm, there is an increase in the solvent accessible surface of the protein. The decreased interactions of the oligosaccharide with the protein surface leads to an increase in the space available to the G0 or G1 (α1,6) oligosaccharide, and exposure of the terminal GlcNAc residue. This suggests a model for G2 and G1(α1,6) in which the sugars are bound to, or have restricted motion relative to, the peptide surface. For G1(α1,3) and G0 glycans terminal β1-2GlcNAc residues become exposed and accessible to lectins such as the *Psathyrella belutina* lectin (Tsuchiya et al 1994). In contrast, examination of the X-ray data suggests that none of the oligosaccharide residues on sugars containing terminal galactose on the α1,6 arm are available for such interactions.

(a) IgG G2:

Approximately 40% of IgG glycans from normal serum IgG are of the G2 type. Based on the X-ray analysis of rabbit Fc, it has been proposed that the available space between the CH2 domains is not sufficient to accommodate G2 sugars on both CH2 domains simultaneously. Interestingly, 24% of IgG in one sample of IgG4, which was glycosylated only in the Fc, contained oligosaccharides in which both the α1,3 and α1,6 arms of each sugar chain were galactosylated (Jefferis et al 1990). It is possible that a change in the conformation of the protein or the sugars may be required to allow the processing enzymes access to develop two fully galactosylated oligosaccharides and to accommodate two G2 glycans in the restricted interstitial space.

(b) IgG G1(α1,3) and G1(α1,6)

The relative proportions of these populations is related to IgG subclass. Analysis of IgG from sub-class specific paraproteins (Jefferis et al 1990) showed that, in all the IgG2 subclass samples analysed and in some of the IgG3 samples, G1(α1,3) predominated over G1(α1,6). In contrast, in polyclonal IgG and in subclasses IgG1 and IgG4 paraproteins, the α1,6 arm was substituted up to four times more frequently than the α1,3 arm. In normal serum IgG glycans

analysed by Parekh (D. Phil. Thesis 1988) the ratio of G1(α1,6):G1(α1,3) was 4:1. These data suggest that the relative affinity of the human Gal β-1,4 transferase (of which only one has been identified to date) for each of the arm specific GlcNAcα1,4 residues may be influenced by differences in the structures of the IgG subclasses. In the ER, IgG is assembled completely with two heavy and two light chains prior to transfer to the Golgi apparatus where complex processing and galactosylation occurs (Bole et al 1986), and processing of the complex glycans takes place in the restricted space which the sugars occupy between the CH2 domains. Subclass differences in the 3-dimensional structure of the protein may variably affect the oligosaccharide-protein interactions which have been shown to be important for stabilising the α1,6 arm. Thus in IgG1 and IgG4 the α1,6 arm terminal GlcNAc may be in a more favourable conformation for interaction with UDP-Gal and β-galactosyl transferase than the α1,3 arm terminal GlcNAc. In contrast, in IgG2 and IgG3 the α1,3 arm displays a higher affinity for the enzyme, suggesting that in these subclasses the presentation of the terminal GlcNAc on the α1,3 arm is more favourable. Two sets of data support the suggestion that the polypeptide structure can influence the branch specificity of the transferase. In the first, Galβ1-4 tranferase from a number of sources was shown to preferentially transfer galactose to the α1,3 arm of symmetrically branched acetylated glycopeptide and oligosaccharide acceptors terminating in GlcNAc (Narashimhan et al 1985). This indicates that either cellular factors or the polypeptide structure influence the branch specificity of the reaction catalysed by the transferase in IgG. Secondly, in fusion constructs of murine myeloma cells, which secrete heterodimeric IgG composed of heavy chain derived from each parental cell, the heavy chains differ in the proportion of each monogalactosylated species. This also suggests that the polypeptide chain can influence *in vivo* the branch specificity of the transferase (Lee et al 1990).

(c) *IgG G0*

The incidence of IgG G0 type oligosaccharides is an age related parameter (Parekh et al 1998a). Overall, G1 does not vary with age, and G2 varies inversely to G0. In patients with rheumatoid arthritis (RA) both serum and synovial IgG contain an increased proportion of glycans which lack galactose and terminate in N-acetylglucosamine (Parekh et al 1984), the levels of which correlate with clinical score and with disease activity (Parekh et al 1988b). 13 pairs of matched RA serum and synovial fluid samples were analysed using the *Psathyrella velutina* lectin (PVL) which interacts with terminal outer arm β1-2 linked N-acetylglucosamine (Tsuchiya et al 1993). In 11 patients PVL binding was significantly elevated ($p=0.007$) in the synovial fluid compared with the paired serum. Elevated levels of IgG0 are associated with a restricted number of other diseases including juvenile RA, tuberculosis, Crohn's disease and systemic lupus erythematosus (Parekh et al 1989).

1.2 Rheumatoid factors

One of the characteristic immunological features of rheumatoid arthritis is the presence in serum and synovial fluid of rheumatoid factors (RFs); however their role in pathogenesis has not yet been established. RFs are defined as antibodies specific to antigenic determinants on

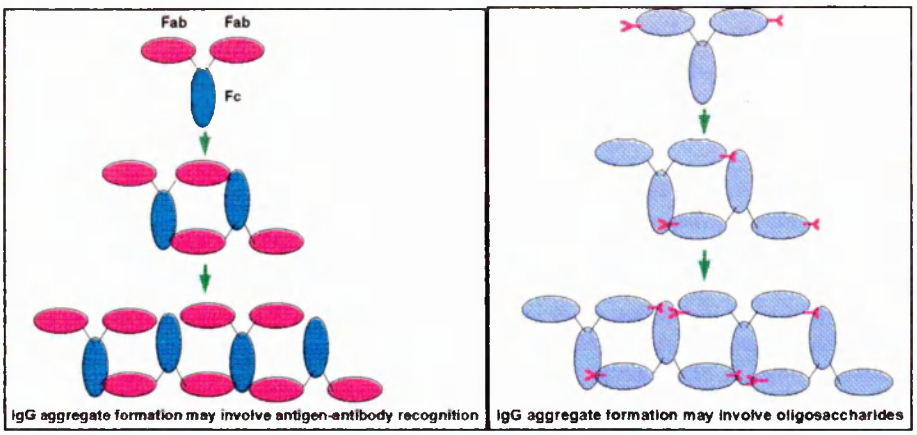


Figure 6a,b: Taken from Nardella et al 1981 and Rademacher et al 1988 respectively

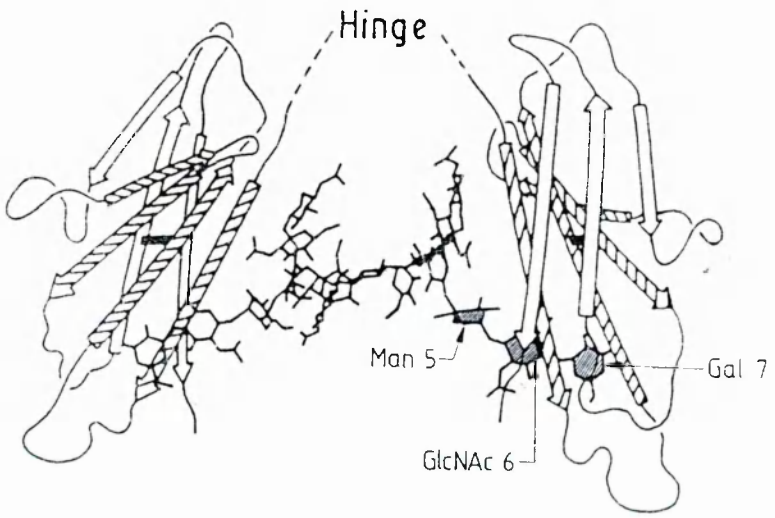


Figure 7a: IgG Fc oligosaccharides

In the IgG molecule the branched oligosaccharide chains cover part of the four stranded faces of the CH2 domains and occupy the space between them. (Taken from Sutton and Phillips 1983).

IgGFc. They exist in three main classes of immunoglobulins, and immune complexes form through the binding of IgM, IgA or IgG to the Fc region of IgG. IgGRFs consist of self associated IgG complexes containing a high proportion of IgG0 glycoforms (85% S-C Chang D. Phil. Thesis). The monomers first form dimers which then self associate to tetramers, hexamers and larger oligomers. IgGRFs may be soluble or deposited on cell surfaces such as the synovium. In synovial fluid they constitute from 1-4% of the total IgG (Leader et al personal communication). The rheumatoid synovium synthesises 12-26% of the IgG in the synovial fluid of RA patients at the rate of 5-95mg/day compared with none in normal synovial fluid (Sliwinski and Zvaifler 1970). The classical model for the structure of IgG-IgG RF complexes is based on antibody-antigen recognition of the Fc of one IgG molecule by the Fab of another (Fig 6a) (Nardella and Mannik 1981). An alternative model for IgG RF complexes has been proposed (Parekh et al 1985) which involves protein-oligosaccharide interactions. This suggests that Fab sugars from one IgG molecule interact with a second through the 'lectin-like' protein surface in the Fc which becomes exposed in the G0 type glycoforms (Fig.6b).

1.3 Protein-oligosaccharide interactions within the IgG molecule

The model of protein-oligosaccharide interactions in Fc has been determined by X-ray crystallography, (Sutton and Phillips, 1983) and the conformational analysis of the free IgG oligosaccharides by high resolution NMR (Rademacher et al 1983, Homans et al 1987). The free sugars contain regions of defined secondary structure and all but one of the glycosidic linkages exist in a single preferred conformation giving rigidity to the core structure. The α1,6 linkage, which can adopt two conformations, gives a flexible arm to the oligosaccharide. In rabbit Fc, the α1,3 arm extends into the interstitial space where the relatively rigid α1,3 arm of one of the oligosaccharides interacts with the core of the opposing sugar while the α1,6 arm of each carbohydrate interacts with the domain surface (Fig. 7a). The crystal structure of human IgGFc predicts a total of 38 non-covalent interactions between the hexasaccharide unit ($GlcNAc_2Man_3GlcNAc$) and the Fc face of the CH_2 domain and additional interactions for GlcNAc (13) and Gal (27) on the α1,6 arm and for the core fucose (7) attached to the primary GlcNAc (Padlan 1991). Some of these interactions are shown in figure 7b.

1.4 The role of IgG Fc oligosaccharides in C1q binding

Protein-oligosaccharide interactions play a role in maintaining the 3-dimensional structure of the two CH2 domains in the Fc. Both aglycosylated and degalactosylated IgG bind less efficiently to C1q and Fcγ receptor (Nose and Wigzell 1983, Leatherbarrow et al 1985, Tsuchiya et al 1989 J. Rheum. 16 285-290). C1q binds to antigen-bound IgG thus activating the classical pathway of the complement system. Partial deglycosylation of IgG does not greatly affect C1q binding (Tsuchiya et al 1989), however complete deglycosylation of IgG does reduce the affinity of C1q, probably due to alterations in the alignment of the IgG CH2 domains (Leatherbarrow et al 1985).

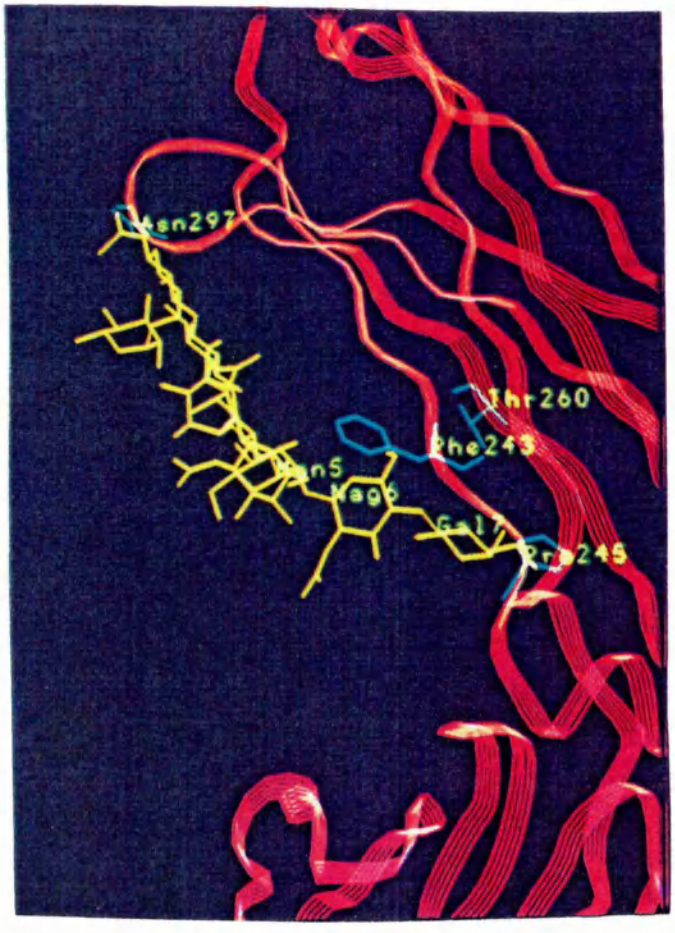


Figure 7b: The CH2 region of the antibody molecule *taken from the crystallographic structure determined by Deisenhofer (1981). The crystal structure of IgG predicts 72 non-covalent interactions between the oligosaccharide and the amino acid residues of the CH2 domain (Padlan 1991). These include interactions between Phe243, Pro245 and Thr260 which are in the proximity of Man5 and GlcNAc6; Gal7; GlcNAc6 and Gal7 respectively of the α1,6 antenna.*

2. Introduction to chapter 6

In general, the analysis of oligosaccharide structures attached to proteins is important both for the characterisation of glycoproteins and for examining the consequences of glycosylation in biological systems. The automation of techniques for releasing and analysing the structures of sugars have highlighted the need for new strategies to profile and sequence oligosaccharides so that the characterisation of glycoproteins can become routine. During the course of this work several new strategies were explored. Conventionally oligosaccharides, released from the protein by hydrazine or with the endoglycosidase PNGaseF, are reduced and radiolabelled at the reducing terminus with tritium. The glycan pool is resolved into its pure components and each is sequenced individually by means of a series of exoglycosidase digestions which sequentially remove the terminal sugars. The products of each incubation are analysed by gel filtration and the partially digested sugar is recovered and subjected to the next enzyme digest. This process has been rationalised by the introduction of the reagent array sequencing method (RAAM) in which the individual oligosaccharides are simultaneously digested with an enzyme array. However the difficulties involved in resolving the glycan pool into its individual components remain. This can be technically difficult when glycans are structurally very similar. Multiple separations are required to separate the glycans, traditionally using P4 GPC to exploit differences in hydrodynamic volume. In this chapter a strategy has been developed in which the need for both labelling and prior separation of glycans has been overcome. In conjunction with Dr. David Harvey, mass spectrometry has been used, first to determine the composition of the IgG glycan pool in a single analysis and subsequently to follow the sequential exoglycosidase digestions of each oligosaccharide simultaneously. HPLC techniques were developed to analyse fluorescent or radiolabelled glycans on the basis of charge, hydrophobicity and hydrodynamic volume. In this chapter WAX chromatography has replaced conventional high voltage electrophoresis (HVE) and cation exchange (CAT) chromatography proved a useful alternative to P4 gel permeation chromatography (GPC). The new technology explored in the course of this project was applied and extended during the work involving the leukocyte antigens in chapter 2.

In this chapter novel strategies were used to analyse the glycans released from normal human serum and RA serum IgG and the results were compared with previous studies by others. IgGRF was prepared to examine its glycosylation and to explore the role of glycosylation in the formation of self associating RF complexes. IgGRF (85% G0) was used in preliminary experiments to test whether the terminal GlcNAc residues on IgG0 may be exposed and available for binding. The results suggested that, if multiply presented on a dot blot, the IgG0 sugars were able to bind MBP. To examine the preliminary finding more rigorously pure sets of IgG0, FabG0 and FcG0 glycoforms were prepared. Normal human serum IgG was cleaved by papain into Fab and Fc, and the natural mixtures of IgG, Fab and Fc glycoforms (containing G2, G1 and G0 glycoforms) were modified to populations containing 100% G0 type sugars by removing sialic acid and galactose residues with *Arthrobacter ureafaciens* and *Streptococcus 6646K* β-galactosidase.

3. Methods

3.1 Preparation of normal and rheumatoid IgG by ammonium sulphate precipitation and DEAE ion exchange chromatography

Whole blood from a rheumatoid patient was incubated at 37°C for 1 hour and at 4°C for 4 hours. The clot was removed and the remaining serum spun for 2 minutes at 2000 rpm to separate it from red blood cells. 1ml of serum was transferred to a centrifuge tube and 500µl of saturated ammonium sulphate pH7.2 were added dropwise to a final concentration of 33%. After incubating on ice for 2 hours the precipitate was recovered by centrifuging for 20 minutes at 5000rpm in a JA21 rotor. The pellet was washed by re-suspending in 40% saturated ammonium sulphate pH7.2 and recovered by centrifugation as before. The pellet was dissolved in 1ml of 20mM potassium phosphate buffer pH7.2, dialysed into this buffer overnight and loaded onto a column of DEAE sephacel (4ml) equilibrated in the same buffer. The unbound fraction contained IgG. The yield was 6.8mg/ml. Normal human serum IgG was prepared by the same method on a large scale from 200ml of normal pooled serum obtained from the blood bank. The yield was 5.2 mg/ml of serum. The preparation was analysed by HPLC, SDS PAGE and western blotting.

3.2-3.5 refer to the methods used for oligosaccharide analysis

The methods used in this chapter and throughout this thesis for oligosaccharide analysis are collected together and discussed in Chapter 8, entitled 'Methods for oligosaccharide analysis'

3.2 Release and labelling of sugars for analysis

The sugars were released from adult and juvenile RA IgG, and from normal IgG, Fab, Fc and their G0 counterparts by hydrazinolysis using the automated OGS GlycoPrep method. The released oligosaccharides were reduced and the reducing terminus radiolabelled or fluorescently labelled.

3.3 Analysis of charged sugars

Fluorescent or radiolabelled sugars were analysed by weak anion exchange (WAX) chromatography.

3.4 Analysis of asialo oligosaccharides

Unlabelled desialylated sugars were desalted by ion exchange chromatography on mixed bed resins prior to analysis by MALDI MS.

3.5 Analysis of the agalactosyl content of IgG

The released and labelled sugars were incubated with the standard 'G0 mix' to remove fucose and N-acetylglucosamine residues. This reduced all glycans to one of three structures differing only in the extent of outer arm galactosylation. The digests were analysed by cation exchange (CAT) chromatography.

3.6 Preparation of monomeric IgG sepharose affinity column

300 mg of globulins (human; Cohn fraction II,III; Sigma) were applied to an HPLC column (TSK-55; 80x2.5cm) equilibrated in 0.05 M potassium phosphate/ 0.15M NaCl pH8. Fractions comprising the monomeric peak were pooled and concentrated (100mg). 90mg of normal monomeric IgG in 60 ml of 0.1M $NaHCO_3$ / 0.5M NaCl pH 8.3 were coupled for 2 hours at 37°C to 11ml of CNBr activated sepharose 4B previously washed in 11mM HCl. After washing with coupling buffer vacant sites were blocked with ethanolamine (1M; pH8; 2h room temperature). The gel was subsequently washed in coupling buffer and in 0.1M sodium acetate/0.5M sodium chloride pH4. It was stored in 0.02M sodium borate/ 0.5M sodium chloride at a final concentration of 7mg of monomeric IgG/ml of gel.

3.7 Sucrose density gradient conditions

A series of sucrose solutions was prepared in which the concentration varied in 5% steps between 10% and 40% sucrose. For a 12ml gradient: 1.77ml of 40% sucrose was frozen at the bottom of a centrifuge tube. 1.77ml of each subsequent dilution (35%, 30%, 25%, 20%, 15%,10%, 5%) was pre-cooled, added to the tube and frozen. Gradients prepared in this way were stored at -20°C until needed. 8-12 hours was allowed for thawing before use. A maximum of 500µl of sample was applied to the top of the gradient and the tube was spun for 22 hours at 4°C and 40K rpm in a Beckman L8-70M using an SW40Ti rotor. Gradients were fractionated using an ISCO density gradient fractionator by pumping a chase solution of 40% sucrose into the bottom of the tube, forcing the gradient out of the top of the tube through a UV detector and into a fraction collector. Immunoglobulins G (7S), A(11s) and M (19S) were well resolved under these conditions.

3.8 Preparation of IgG oligomer standards by chemical cross linking of IgG

Human IgG (800µl; 58nmoles; 11mg/ml potassium phosphate buffer 0.1M, pH7.2) was dialysed into sodium borate buffer (0.1M, pH9). An equimolar quantity of dithiobis-succinimidyl proprionate (DTSP) (23.7µg/5.9µl DMF) was incubated in a reactivial for 7 days during which time 3 further aliquots of DTSP/DMF were added. The extent of cross linking was monitored by HPLC (TSK 3000) equilibrated in 0.1M potassium phosphate buffer pH7.2, and the final equilibrium mixture fractionated by SDG 0-40%).

3.9 Preparation of IgGRF and RF depleted IgG

Serum (3ml) from a patient with hyperviscosity syndrome and RA (G0 45%) was applied to a column of CNBr-activated sepharose 4B (13ml) to which 90mg of normal monomeric IgG had been coupled. The column was equilibrated and washed with 0.02M disodium tetraborate/0.5M sodium chloride pH 8.8. Bound material was eluted with 0.1M sodium acetate/0.15M sodium chloride pH 3.5, pooled, concentrated to approximately 7mg/ml and fractionated by sucrose density gradient (12ml; 10-40%, pH3.5 SDG; centrifugation: 16h, 52k, 4°C). At pH 3.5 IgGRF is

dissociated from IgG (7s), IgM (19s) and IgA (11s). The 7s pool containing IgGRF monomers was concentrated and dialysed into 0.1M glycine/0.16M sodium chloride pH7. Sedimentation values were assigned by comparison with known standards obtained by chemically cross-linking IgG and from standard IgG and IgM. Unbound material from the monomeric IgG sepharose column, which contained IgG depleted of RF, was concentrated to the original serum volume (3ml) and IgG prepared as in 3.1.

3.10 Papain cleavage of IgG to investigate whether the % uncleaved IgG correlates with G0

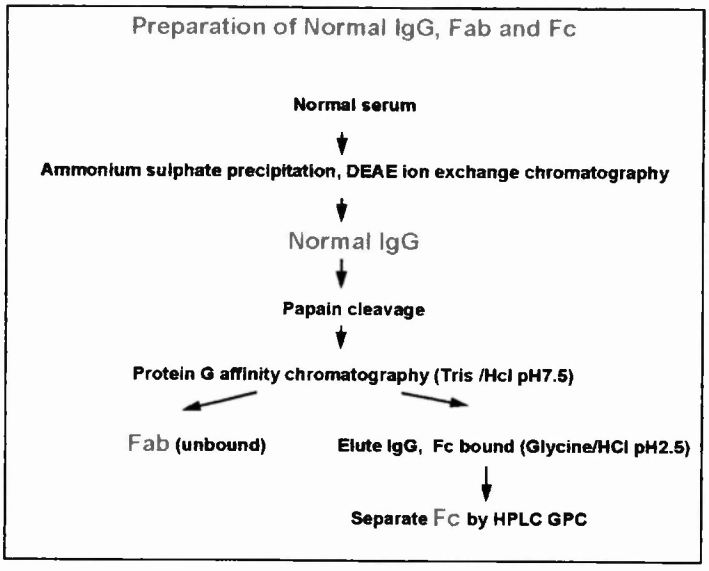
Human serum IgG was isolated from 20 juvenile RA patients by ammonium sulphate precipitation and subjected to papain cleavage. IgG (50μg) in 20mM potassium phosphate pH7 (5mg/ml) was incubated at 37°C for 4 hours with 0.1ml of sepharose CL4B derivatised with papain (Pharmacia) in 300μl of buffer (20mM potassium phosphate pH7/Cys/HCl 20mM/EDTA 4mM; enzyme substrate ratio 1:40). The percentage of papain resistant IgG was calculated by integration of the IgG, Fab and Fc peaks from analytical GPC (Zorbax GF250+450). The percentage of papain resistant IgG was plotted against the G0 values for each IgG sample.

3.11 Papain cleavage of IgG to prepare Fab and Fc

IgG (500μg in 500 μl 20mM potassium phosphate pH7) was incubated at 37°C with 1ml of sepharose CL4B derivatised with papain (Pharmacia) in 3ml of buffer (20mM potassium phosphate pH7/Cys/HCl 20mM/EDTA 4mM; enzyme substrate ratio 1:40). The reaction was followed by HPLC and reached equilibrium after 4 hours. After extensive dialysis into Tris/HCl pH7.5 (10mM) Fab was separated from IgG and Fc by Protein G affinity chromatography.

3.12 Protein G sepharose affinity chromatography to separate IgG and IgG Fc from IgG Fab

The sample was dialysed into Tris/HCl (10mM; pH7.5) and applied to a protein G affinity column (5x5cm, selectisphere 10 Perstorp Biolytica). The flow rate was 0.1ml/min and absorbance was monitored by UV detection at 280nM. The column was washed extensively to remove unbound Fab. IgG and Fc were eluted with 100mM glycine/HCl pH2.5. The pH of the fractions was immediately adjusted to 7 with 1M Tris to protect sialic acid residues. Both bound and unbound fractions were passed through the column a second time to remove any residual Fab from the IgG/Fc and vice-versa. Fc was separated from uncleaved IgG and any low molecular weight minor products which might have resulted from the papain cleavage by HPLC GPC. Fab, which does not contain additional papain cleavage sites, was concentrated in centricon filters (10k cut off membrane) and the buffer partially exchanged for sodium acetate pH5.5/10mM $MnCl_2$ before finally dialysing into the same buffer. Following HPLC purification Fc fractions were concentrated and dialysed in the same way.


Preparation of Normal IgG, Fab and Fc
Normal serum
Ammonium sulphate precipitation, DEAE ion exchange chromatography
Normal IgG
Papain cleavage
Protein G affinity chromatography (Tris /Hcl pH7.5)
Fab (unbound)
Elute IgG, Fc bound (Glycine/HCl pH2.5)
Separate Fc by HPLC GPC

Flow Chart 1

3.13 De-galactosylation of IgG, Fab and Fc by Arthrobacter ureafaciens neuraminidase and Streptococcus 6646K β-galactosidase

Normal serum IgG (800μg), Fab (800μg) and Fc (200μg) in 0.05M sodium acetate pH 5.5 containing 10mM $MnCl_2$ at a concentration of 1mg/ml were incubated with *Arthrobacter ureafaciens* neuraminidase (2U/ml) and *Streptococcus* 6646K β-galactosidase (0.2U/ml) (Siekagaku Co. Ltd.) at 40°C for 16h. The method used was a modification of the conditions used in Kiyohara et al 1976. The reactions were terminated by incubating at 56°C for 10 minutes. Control samples were treated in the same manner without the addition of enzymes. IgG, Fab and Fc sugars were analysed before and after (Kiyohara et al 1976) enzymatic removal of terminal GlcNAc and fucose. The glycans were released by hydrazine and labelled either with tritium or 2AB. Labelled glycans were analysed according to charge using WAX chromatography to confirm the digestion of sialic acid residues in the enzymatically modified samples. Following incubation of the glycans with the 'G0 mix' (ch.8, section 1.4h) the %G0 of the modified IgG, Fab and Fc was determined by cation exchange chromatography (ch.8, section 2.3: Vydac 400 VHP575 column; gradient 0-80% MeCN containing 50mM formic acid) .

3.14 Detection of IgG binding to MBP in a preliminary study

A plate assay to detemine IgG0 developed by Dr. Graham Rook (personal communication) was modified for use in a dot blot assay. 5μl of each IgG solution were spotted on to nitrocellulose blotting membrane impressed with a dot matrix by means of the Biodot apparatus (Biorad). The membrane was blocked with 5% BSA, 0.05% Tween in PBS for 16 hours at room temperature. After washing with 2x blocking solution and 1x PBS the protein was fixed with 0.5% glutaraldehyde / PBS for 30 minutes at 4°C. The membrane was washed with 0.1% lysine /PBS, incubated at 37°C for 90 minutes and for 16 hours at 4°C with MBP (10μg/ml/Ca^{2+}/PBS). The membrane was washed as before and incubated with a polyclonal rabbit IgG α–human MBP. After a final wash the bound antibody to MBP was detected by incubating with rabbit IgG conjugated to alkaline phosphate for 90 minutes at room temperature. Visualisation was with NBT/BCIP.

4. Results:

4.1 Preparation and analysis of normal and rheumatoid IgG (Flow chart 1)

Normal and RA IgG were prepared by ammonium sulphate precipitation and DEAE anion exchange chromatography. The yields were 5.2 mg/ml and 6.8 mg/ml of serum respectively. The purity of the IgG was assessed by HPLC (Fig. 8) and SDS PAGE (Fig. 9, tracks 1 and 2). IgG heavy chain was detected at approximately 50kD and IgG κ and λ light chains at approximately 32Kd and 29 kD respectively. No μ- (70kD) or α- (55kD)chain was detected and the absence of IgA and IgM was confirmed by Western blot analysis. Another possible contaminant is the C1 complex which may be bound to IgG and therefore co-isolate. No bands were visible on the gel at 22, 23, 24 or 83kD, the molecular weights of the components of the

C1 complex after reduction. However such contaminants may be present at levels of less than 1% of the total protein.

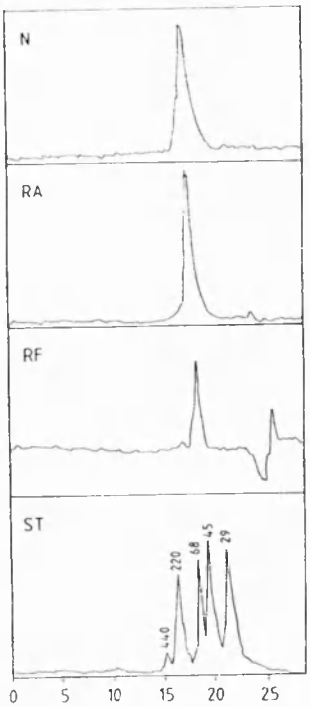


Figure 8: *HPLC GPC analysis of (i) normal and (ii) RA IgG (iii) IgGRF (iv) standard proteins using a tandem column consisting of Zorbax GF 250 + 450 Zr-silica columns (2x9.4x250mm). The buffer was 0.1M potassium phosphate pH7.2.*

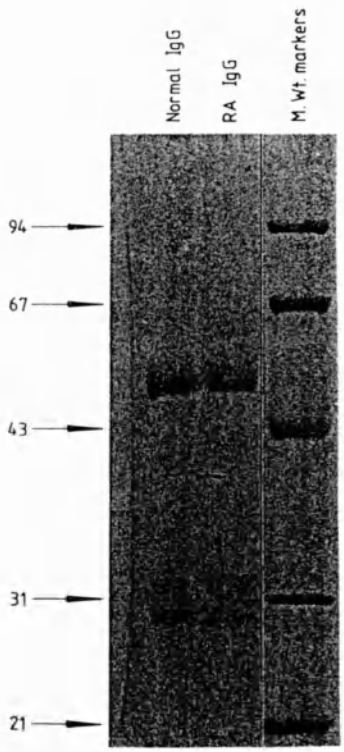


Figure 9: *SDS PAGE (10% reducing gel; Coomassie stained) of normal (track 1; 10μg) and rheumatoid (track 2: 5μg) IgG. The heavy chains have an apparent molecular weight of approximately 50kD and the light chains approximately 29 (κ) and 33 (λ) kD. Track 3 contains low molecular weight markers.*

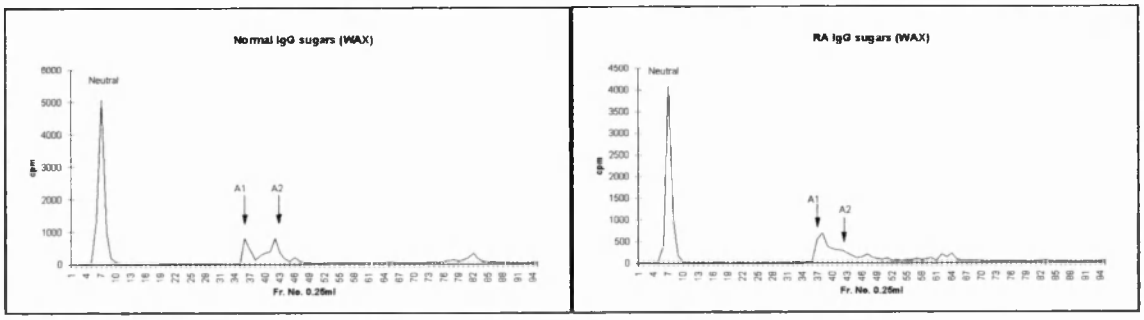
4.2 Glycosylation analysis of normal and rheumatoid IgG

4.2.1 Release and labelling of IgG glycans

Sugars were released from 1mg of normal and 1mg rheumatoid IgG by hydrazinolysis using the Oxford GlycoSystems (OGS) GlycoPrep 1000 (85% yield (OGS specification; 16.6nmoles). 80% of the sugars were labelled with tritium (13.3nmoles, approximately 1.7E8 counts), and 20% retained unlabelled. 50% (6.6nmoles) of each labelled glycan pools was desialylated with *Arthrobacter Ureafaciens* sialidase.

4.2.2 Analysis by charge of radiolabelled glycans

A charge profile of both the normal and RA IgG radiolabelled glycan pools was obtained by weak anion exchange chromatography (WAX) (Fig. 10a,b). Aliquots containing 1E3 counts were injected and the profiles obtained by handcounting the radioactivity in the eluted fractions. In both cases more than 75% of the structures were neutral, and all structures became neutral after incubation with *Arthrobacter ureafaciens* neuraminidase (data not shown) indicating that all charge was due to sialic acid.



76% neutral; 14% A1; 10%A2

80% neutral; 15%A1; 5%A2

Figure 10a,b Charge analysis of tritium labelled normal and RA IgG sugars by WAX *After incubation with* Arthrobacter ureafaciens *neuraminidase all the structures become neutral, indicating that all charge is due to sialic acid. The profiles were obtained by hand counting the fractions eluted from the chromatography column. The flow rate was 1ml/min and 0.25ml fractions were collected. The gradient was applied at 12 minutes.*

4.2.3 Analysis by charge of 2AB labelled glycans using on line fluorescent detection

A charge profile of both the normal and RA IgG fluorescent labelled glycan pools was obtained by weak anion exchange chromatography (WAX) (Fig. 10c,d). The elution was monitored using an on line detector. The samples were applied to the column through an automatic injector which accounts for the broadening of the neutral peak. The gradient was applied at 12 minutes as in figure 10a,b.

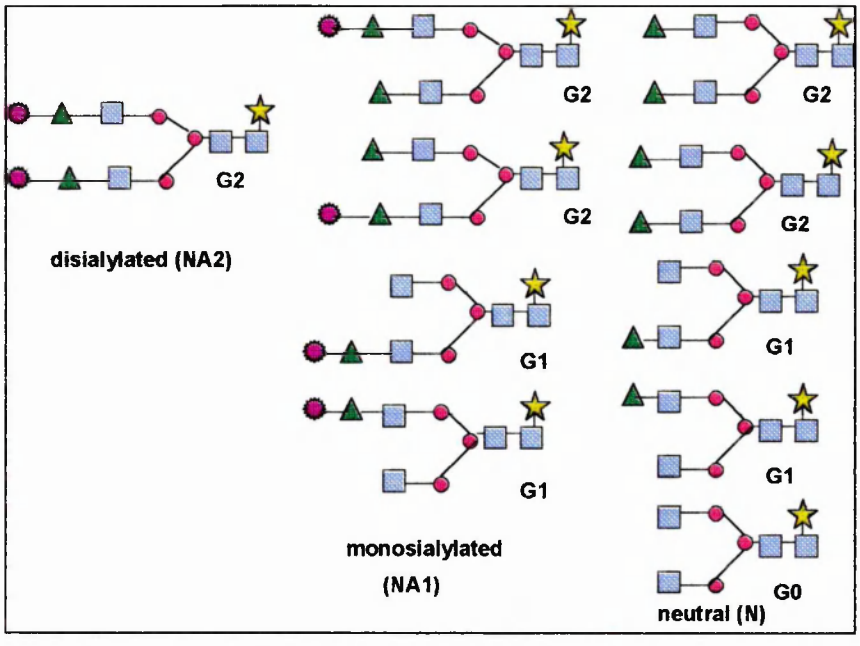


Figure 10e: Schematic illustration of the sets of charged glycoforms present in IgG. *This refers to figure 10c,d. The core structure shown contains fucose, other core types (see figure 5a) are also present (see Table 1). RA IgG contains both the G1 and G2 sets of monosialylated A1 type structures, while normal IgG contained only the G1 set.*

The improved technology indicates that the RF monosialylated (A1) population contains two structures. By comparison with standards the front of the peak was identified as the pool of G1 sugars in which the single terminal galactose is sialylated while the late shoulder contained G2 sugars in which only one of the terminal galactose residues has been sialylated. In contrast the data for normal IgG glycans indicate that this glycan pool contain only monosialylated G1, and that the G2 glycans are either disialylated (A2) or neutral (N).

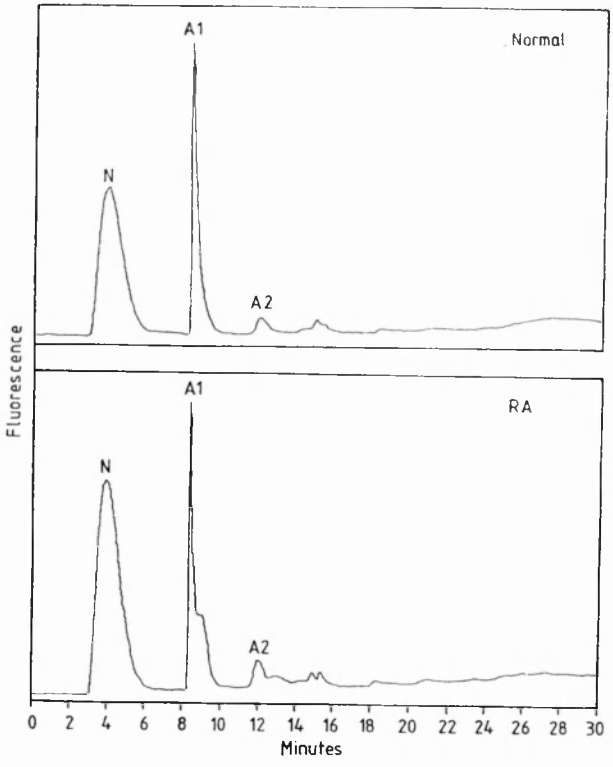


Fig. 10c,d Analysis of 2AB labelled normal and RA IgG sugars by WAX chromatography *with on line detection. The samples are identical to those in figures 8 and 9 except that the sugars have been labelled with 2AB instead of tritium.*

4.2.4 Analysis of asialo IgG glycans by size using P4 GPC

De-sialylated tritium labelled IgG sugars from normal and rheumatoid IgG were profiled by P4 GPC (Fig. 11 a,b). In each case fractions containing 1E5 cpm of radioactivity were applied to the column. The samples were co-injected with an internal standard containing a mixture of glucose oligomers prepared from a dextran hydrolysate. The elution of the glucose oligomers was monitored by refractive index (RI) while the samples were monitored by a radioactivity detector. The hydrodynamic volumes of the peak fractions which are marked were obtained by overlaying the RI and radioactive traces.

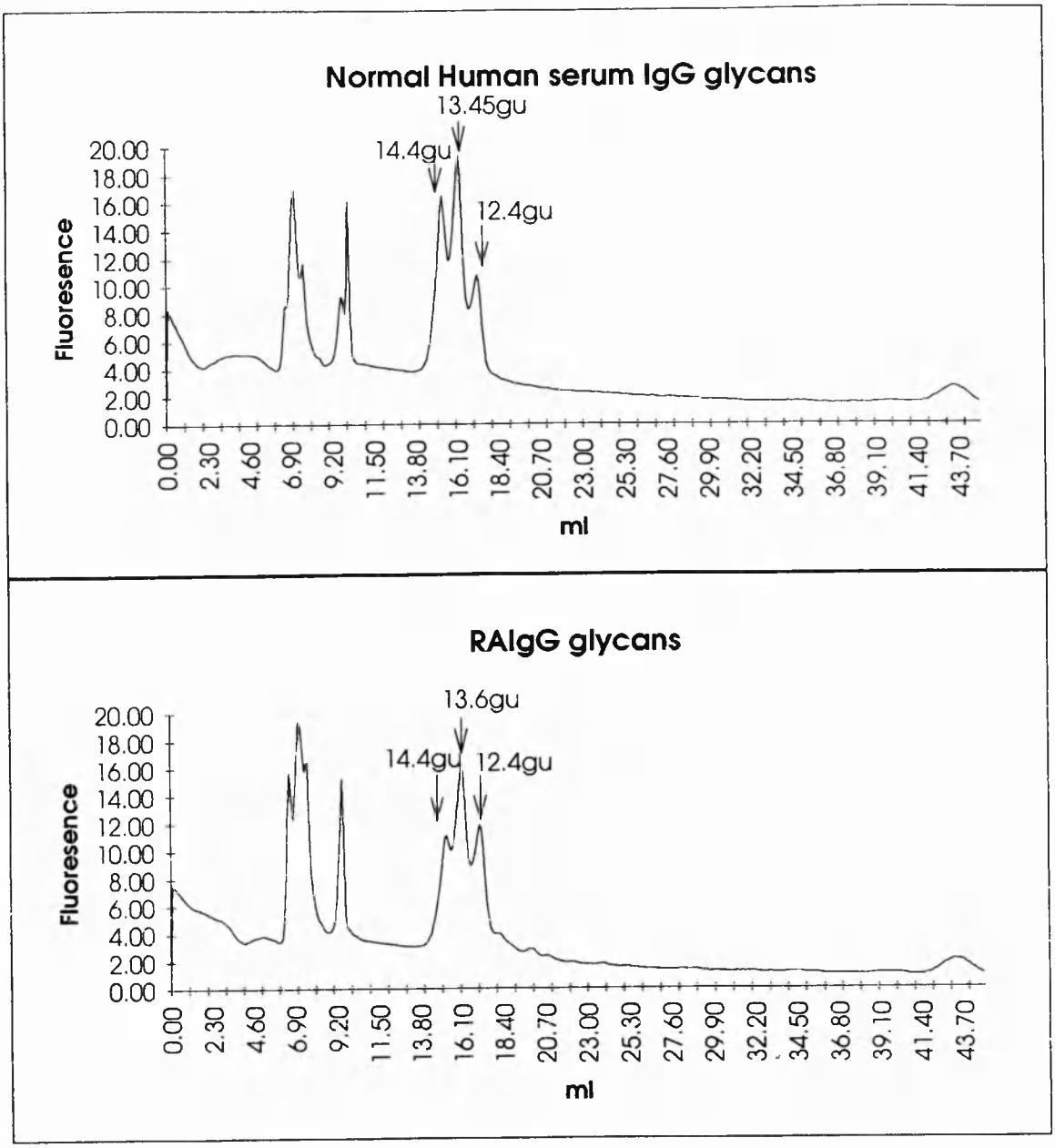


Figure 11 c,d: P4 GPC profiles of 2AB labelled asialo glycans from IgG from normal and RA serum. These samples are from the same glycan pool as those analysed in 11a,b. The enhanced resolution is a result of the fluorescent labelling of the glycans which each elute within a shorter time period than the corresponding tritium labelled samples, allowing better resolution of the different species. The material eluting before and in the void (10ml) is an artefact of fluorescent labelling, the origin of which is currently unknown.

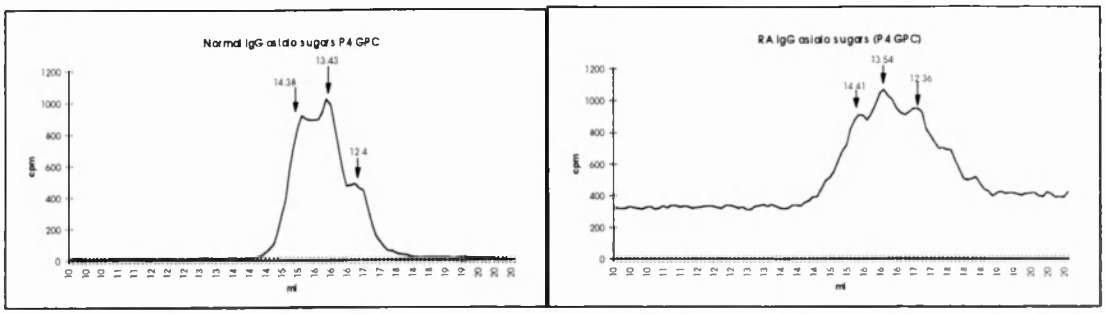


Figure 11 a,b: P4 GPC GlycoMap profiles of tritium labelled asialo glycans from normal and RA IgG

11c,d (opposite) show the P4 analysis of the same samples fluorescently labelled.

At this point the conventional sequencing strategy requires the individual species within these envelopes of peaks to be separated into individual components which are at least 80% pure. Reagent arrays could then be used to sequence each sugar in turn. The alternative approach developed in this chapter used the unlabelled glycan pool released by hydrazinolysis. This was desialylated with *Arthrobacter ureafaciens* sialidase, desalted on mixed bed ion exchange resins and analysed without further fractionation by MALDI MS.

4.2.5 Analysis of unlabelled asialo glycans from normal IgG by MALDI MS.

16 oligosaccharides were identified by MALDI MS (Fig.12) beneath the incompletely resolved envelope of structures in fig. 11a. and these are shown in Table 1. In Table I the population of glycans identified in this MS analysis has been compared with the set of glycans released from polyclonal IgG (Cohn Fraction II) and resolved by HPLC (Jefferis et al 1990). This contained 15 asialo glycans of different molecular weights. The composition of the glycans released from normal serum IgG and analysed by P4 GPC (Parekh D.Phil. Thesis 1987) is also shown for comparison. In this case 11 asialo glycans of different molecular weights were identified. The three sets of data, which relate to different samples of polyclonal IgG analysed by different analytical techniques, are in good agreement. The increased resolution obtained by HPLC reveals three of the four members a set of structures not previously reported in which the α1,6 arm of the biantennary glycan terminates with mannose. These structures were identified by MALDI MS and also the fourth (H3N3) which is the unfucosylated form of H3N3F reported by Jefferis. Although these data are consistent with those obtained by Jefferis et al it is possible that some of the minor components (those present at <2%) of the mixture of glycans analysed in this chapter may be derived from contaminants in the IgG preparation or introduced from the environment. The purity of the IgG was assessed by SDS PAGE (Fig.9), and by Western blot analysis. There was no detectable μ- or α- chain to suggest the presence of IgA or IgM contamination nor of the components of the C1 complex. Such contaminants, if present, would constitute < 1% of the total protein. In the MS profile there are unidentified peaks with molecular weights which cannot be assigned to oligosaccharides, suggesting that there is some contamination of the sample. It is also possible that the structures A3, B3, C3 and D3, which contain a bisecting GlcNAc residue but not a core fucose residue, are derived from a

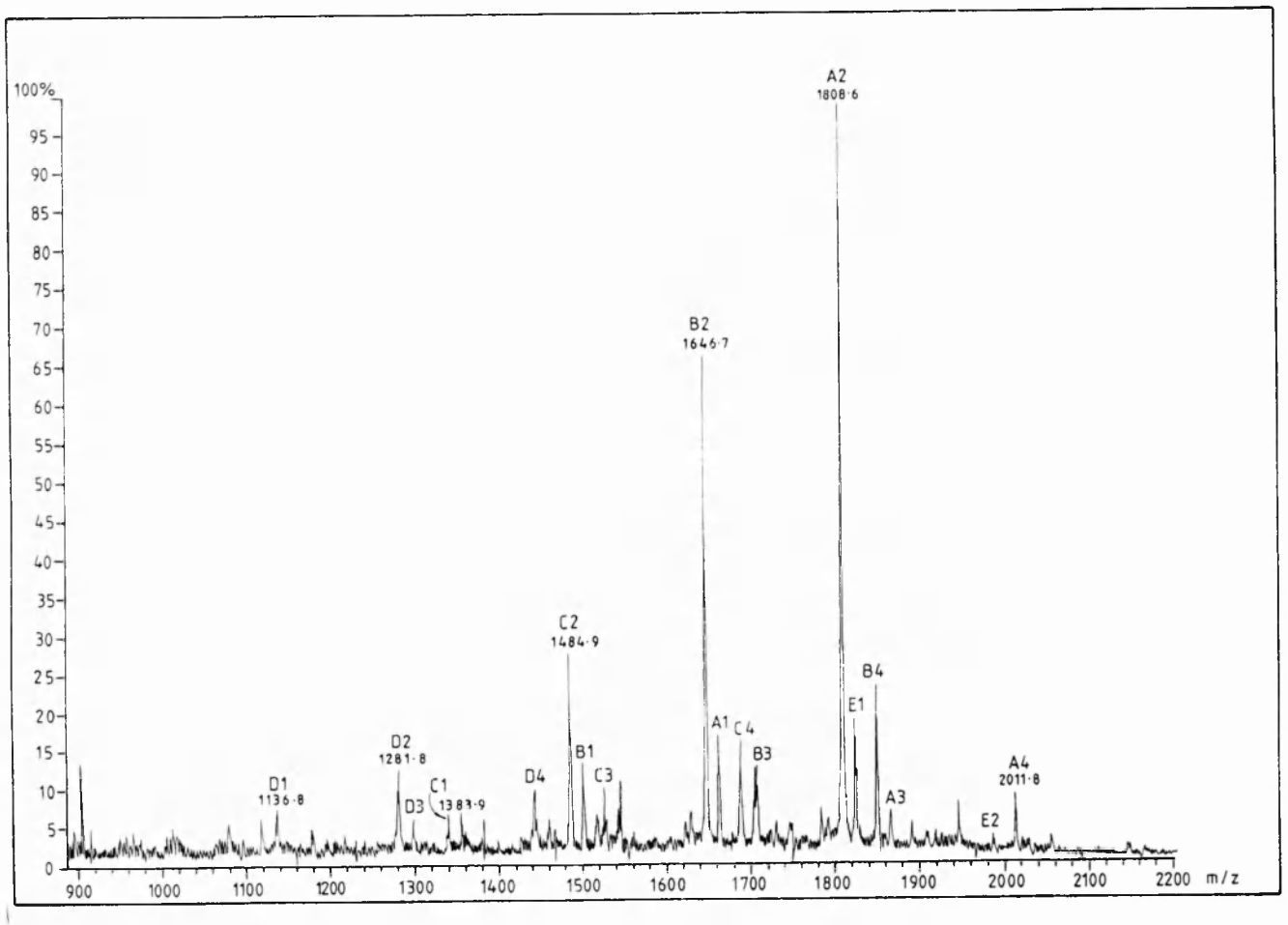


Figure 12: Unlabelled asialo sugars from normal IgG analysed by Matrix Assisted Laser Desorption Mass Spectrometry (MALDI MS). *The data shown is compiled from 25 shots. It is analysed in Table 1.*

hydrazinolysis artefact in which the reducing terminal is cleaved (Mr. Tom Naven -personal communication). In this analysis by MALDI MS the percentage of each structure was derived from measurements of peak height. This correlation should be treated with some caution. Although it is in general reliable, and these results are consistent with the findings of others, it should be noted that it has not been demonstrated that the efficiency of ionisation is identical for every oligosaccharide.

Pk	M.Wt.	Exp.M. Wt.	%	% Jefferis	% Parekh	Formula	Glycosylation
A1	1663.6	1662.6	5.5	4.7	8.2	H5N4	G2 - F-B
A2	1809.7	1808.6	32.5	24.6	22.3	H5N4F	G2+F
A3	1886.7	1865.6	1.5	2.3	1.1	H5N5	G2+B
A4	2012.7	2011.8	2.2	6.2	6.3	H5N5F	G2 + B+F
B1	1501.5	1500.7	3.6	4.8	9.1	H4N4	G1- F-B
B2	1647.6	1646.7	20.9	24.8	24.5	H4N4F	G1+F
B3	1704.6	1703.7	3.5	1.8	-	H4N5	G1+B
B4	1850.7	1849.7	7.2	6.7	1.6	H4N5F	G1 + B+F
C1	1339.5	1338.9	0.9	2.6	2.8	H3N4	G0 - F-B
C2	1485.5	1484.9	8.8	13.9	18.7	H3N4F	G0 +F
C3	1542.7	1525.7	2.2	1.3	1.0	H3N5	G0+B
C4	1688.6	1687.9	4.4	3.2	1.2	H3N5F	G0 + B+F
D1	1136.4	1136.8	1.1	ND	ND	H3N3	GN1*
D2	1282.5	1281.8	2.2	1.3	ND	H3N3F	GN1+F
D3	1298.5	1297.8	0.9	0.6	ND	H4N3	G1-GN
D4	1444.5	1443.8	2.6	1.2	ND	H4N3F	G1-GN+F

* Not previously reported ND - not detected

Table 1a: MALDI MS analysis of sugars released by hydrazine from normal IgG

The structural formulae correspond to the molecular weights and are without arm specificity. The glycans have been classified according to the extent of outer arm galactosylation. On this basis there are three types of asialo biantennary sugars: those which contain a galactose residue on each arm (G2), those which contain one, either on the α1,3 or α1,6 arm (G1(α1,3) and G1(α1,6)) and those which contain no galactose (G0) (Fig. 5). The percentage of each glycan structure was derived from measurements of the peak heights. The percentages obtained by Jefferis et al (1990) for the analysis of polyclonal human IgG sugars are given for comparison. The percentages of G2:G1:G0 were 41.7%:38.7%:19.6% compared with

Glycan	Exp. M.Wt	BTgalact	JB hexos	CL fucosid
A1	1662.6	1338.8	933.5	933.2
A2	1808.6	1484.9	1079	933.2
A3	1865.6			
A4	2011.8	1688.8	933.5	933.2
B1	1500.7	1338.8	1079	933.2
B2	1646.7	1484.9	1079	933.2
B3	1703.7			
B4	1849.7	1688.8	1079	933.2
C1	1338.9	1338.8	933.5	933.2
C2	1484.9	1484.9	1079	933.2
C3	1541.8			
C4	1687.9	1688.8	1079	933.2
D1	1136.8	1136.8	933.5	933.2
D2	1281.8	1281.8	1079	933.2
D3	1297.8	1135.7	933.5	933.2
D4	1443.8	1281.7	1079	933.2

Table 1b : Peaks identified on Figure 12 (a-d)

The species A3, B3 and C3 were not detected after the digestion with β-galactosidase, and the D series were present in vanishingly small amounts. This may be due to sample loss and to the increased proportion of the other peaks by comparison.

41%:39%:20% obtained using the G0 mix and cation exchange chromatography (Fig.15a) (Jefferis: 37.8%:39.9%:22.3%). 80.6% of IgG sugars are fucosylated and 24.5% are bisected (Table 1a). (Jefferis: 81.9% fucosylated, 21.5% bisected.) It was noted that two other series appear in this profile, but in very low abundance. These are a series of oligomannose structures and a series of compounds with molecular weights compatible with a series of hybrid glycans. Such species have been detected previously in IgG (Harada et al 1987).

The experimental molecular weights shown are those determined in Figure 11a and correspond to the structural formulae shown in table 1a. The theoretical molecular weights are the monoisotopic molecular weights calculated as discussed in the chapter 8 section 2.1). This discrepancy indicates that, in this instance, the instrument has not been accurately calibrated. Different oligosaccharides have characteristic molecular weights; however those which differ only in arm specificity cannot be distinguished by this technique, neither is any linkage information directly available in the spectrum. The structures predicted from the analysis of the glycan pool were confirmed by simultaneous exoglycosidase sequencing. This novel approach to sequencing was developed to avoid the problems associated with separating the pool of released IgG glycans into individual pure sugars. It was made possible by the recent developments in this laboratory in the application of mass spectrometry to the analysis of oligosaccharides.

(iv) Simultaneous exoglycosidase sequencing of the pool of asialo normal IgG sugars analysed by MALDI MS in conjunction with Dr. D. Harvey

The total pool of unlabelled asialo glycans was successively digested with Bovine testes β-galactosidase, Jack bean β–hexoseaminidase and *Choronia lampas* α-fucosidase. The scheme is shown in Fig. 13 and the data in Fig. 14 and table 1b. The glycans in the pool have been simultaneously exposed to each enzyme in turn. Analysis of the profiles allows the effect of a single enzyme on each of the major glycan populations to be assessed after each digestion.

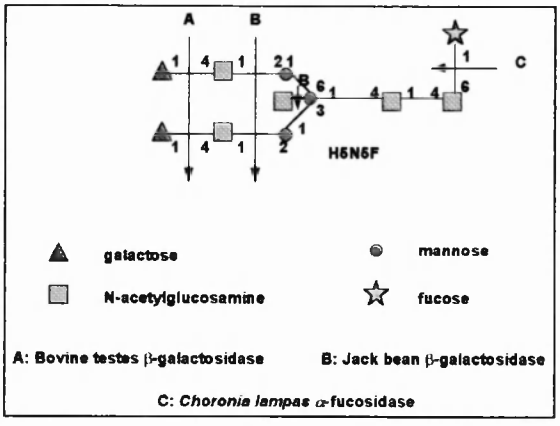


Figure 13: Schematic diagram showing the cleavage positions of the enzymes used to sequence simultaneously the IgG asialo glycans

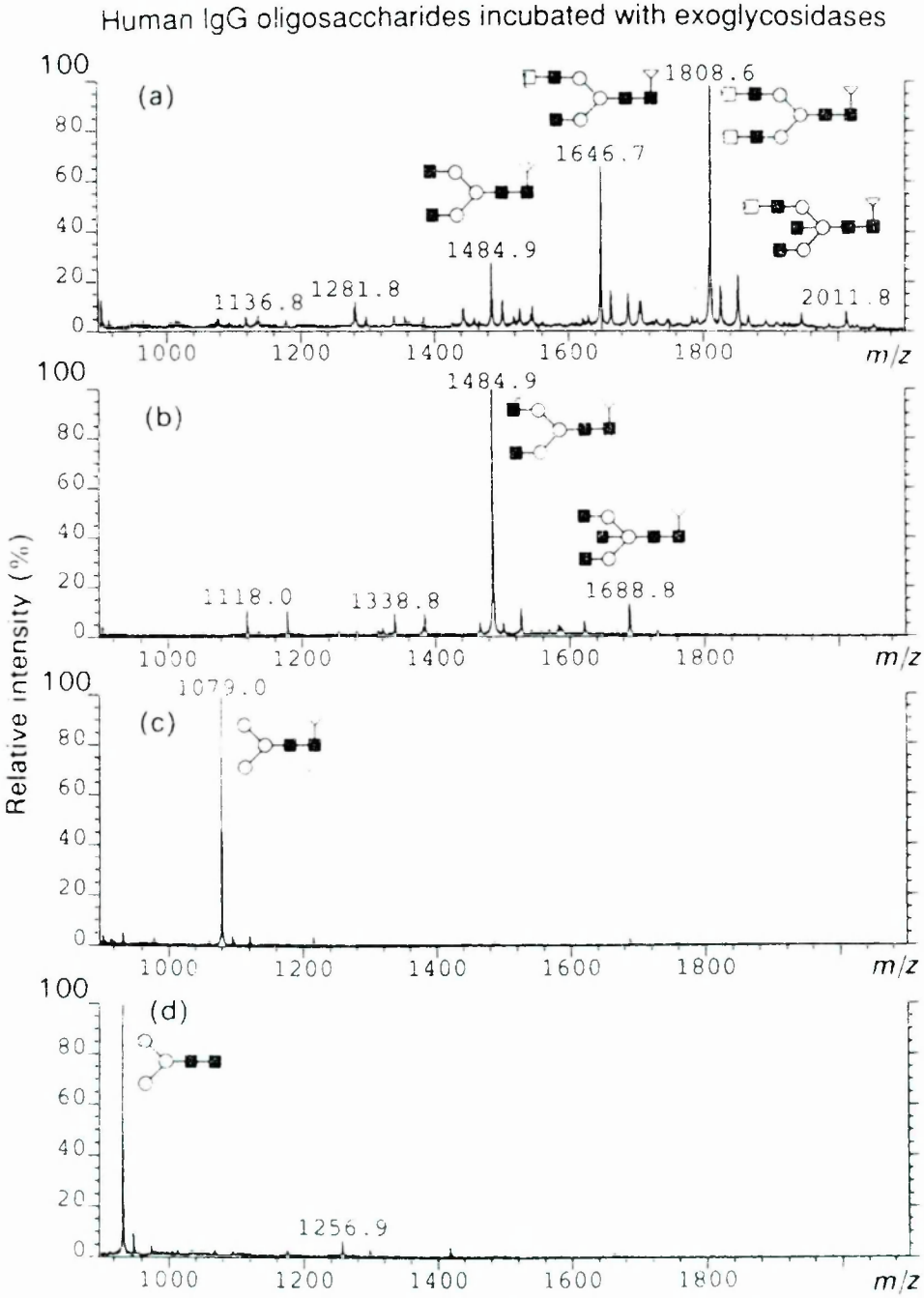


Figure 14. MALDI mass spectra of the products of successive exoglycosidase digests of the *N*-linked oligosaccharides released from human IgG by hydrazinolysis. The matrix in each case was 2,5-DHB. (a) Total oligosaccharide fraction; (b) incubation with bovine testis β-galactosidase; (c) incubation with jack beam β-hexoseaminidase; (d) incubation with *Charonia lampas* fucosidase.

Digestion A: *Incubation with Bovine testes β-galactosidase removed all terminal galactose residues with a loss in molecular mass of 162D for each hexose residue digested.* Digestion B: *Incubation of the products from A with Jack bean β–hexoseaminidase removed terminal N-acetylglucosamine residues.* Digestion C: *Incubation of the products from B with Choronia lampas α-fucosidase removed the core fucose residues. After each digestion the products were desalted and analysed by MALDI MS (Fig. 14a-d).*

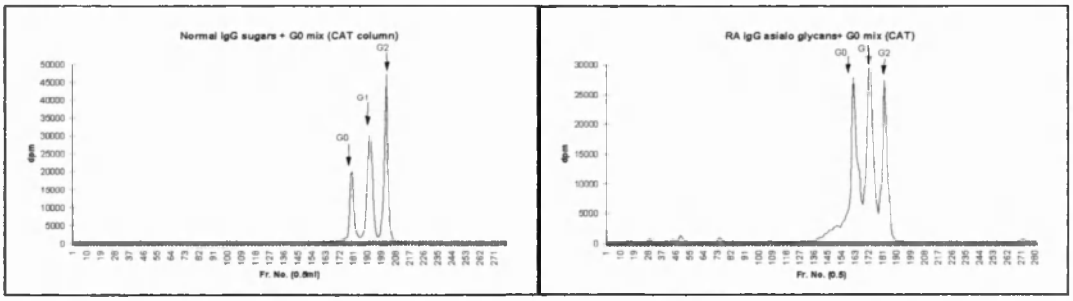
The pool of IgG glycans was first analysed by MALDI MS (Fig. 14a) and then successively digested with Bovine testes β-galactosidase (Fig. 14b), Jack bean β–hexoseaminidase (fig.14c) and *Choronia lampas* α-fucosidase (Fig. 14d). After each digestion the samples were desalted on mixed bed resins and analysed by MS, the analytical details are in table 1b. A comparison of figures 14a and 14b shows that the peak at 1808.6 mass units has been reduced to 1489.4 mass units by Bovine testes β-galactosidase which has removed two galactose residues. Simultaneously the peak at 1646.7 has been digested to 1484.9 mass units by the removal of one galactose residue. The minor peak at 1849 containing a bisecting N-acetylglucosamine residue has been digested to a mass of 1688.8 mass units by the loss of the single outer arm galactose. Other minor structures have all been reduced in size by amounts equivalent to the number of terminal galactose residues they contain. In the next digestion with Jack bean β–hexoseaminidase (14c) all structures have been reduced by the number of terminal GlcNAc residues they contain; for example, the peak at 1484.9 in figure 14b is now detected at 1079 mass units having lost two GlcNAc residues. In the final digestion (fig. 14d) core fucose was removed from the predominant structure in 14c. It can be seen that, as the digestions proceed, the multiple oligosaccharides present in the original pool are gradually reduced to a single common structure, the trimannosyl core. These enzyme digestions confirm the assignments in figure 14a which were based on molecular weight.

4.2.6 Analysis of the G0, G1 and G2 glycoform populations of normal and RA serum IgG.

The percentages of G2: G1: G0 derived from the MS data were 41.7 : 38.7% : 19.6% respectively. The uncertainty about the validity of correlating peak height with the quantity of glycan in the sample indicated that these parameters should also be determined by converting the glycan pool to 3 populations distinguished only by their outer arm galactosylation. An enzyme mix and incubation conditions were optimised in conjunction with OGS for this purpose. The pool of IgG glycans released by hydrazine, radiolabelled and desialylated with *Arthrobacter ureafaciens* neuraminidase was incubated for 16 hours at 37°C with a standard G0 mix containing Jack Bean β-hexoseaminidase and Bovine epididymus α-fucosidase. D-galactonic acid γ-lactone was added to inhibit the activity of β-galactosidase which may be present as an impurity in the Jack Bean β-hexoseaminidase. (Ch. 8 section 1.4h). The enzymes remove outer arm and bisecting GlcNAc residues and core fucose thus reducing all the glycoforms of IgG to three populations which differ only in the extent of terminal galactosylation. After desalting the digest was analysed by cation exchange chromatography

(Fig. 15 a) and the quantities of G2, G1 and G0 obtained by integration of the appropriate peaks.

Normal IgG contained the following percentages of differently galactosylated sugars, G2: 41%, G1: 39%, G0: 20%. These are in good agreement with those obtained by MS: G2:41.7%, G1: 38.7% and G0: 19.6%. RA IgG glycans were also analysed using the G0 mix (Fig. 15b); the corresponding values were: G2: 28%, G1: 36% and G0: 36%.



(a) G0: 20%, G1: 39%, G2: 41% (b) G0: 36%, G1: 36%, G2: 28%

Figure 15 a,b: Cation exchange chromatography of asialo IgG glycans digested with a mixture of fucosidase and hexosaminidase (G0 mix). *The profiles were obtained by hand counting the tritium labelled fractions from the column. The peaks were assigned by comparison with pure samples of G2, G1 and G0 glycans. The percentages of G0, G1 and G2 were obtained by integration of the corresponding peaks.*

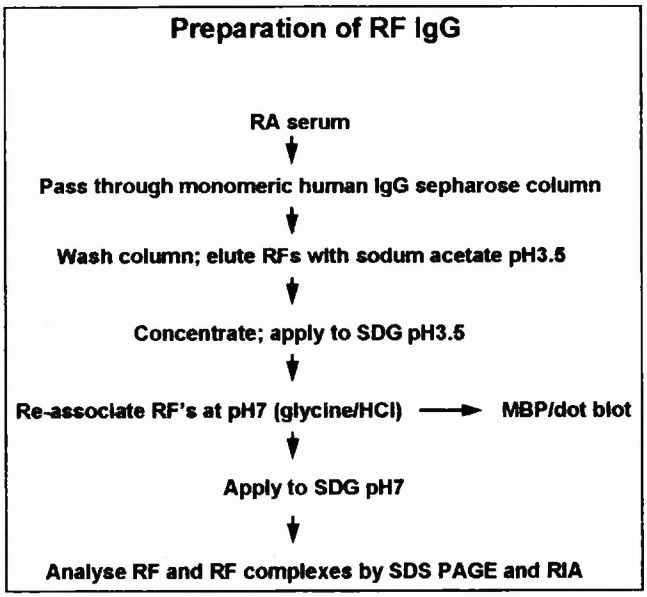
The oligosaccharide analysis of a subset of IgG molecules

The total serum IgG from RA patients has been shown to contain increased levels of IgG0 compared with normal controls. Another characteristic feature of RA is the presence in serum and synovial fluid of IgG rheumatoid factors (IgGRFs). IgGRFs are a subset of IgG molecules which bind IgA, IgM and also IgG and are expressed by particular clones of B-cells. The glycosylation of IgGRF has not been determined and it is not known whether, in addition to possessing common protein features in the variable region, IgGRFs also constitute a specific subset of IgG glycoforms. In the following section therefore, self associating IgGRFs were prepared from RA serum and the glycosylation was examined.

4.3 The preparation and glycosylation analysis of IgG RF

4.3.1 Preparation of a standard IgG oligomer preparation

This was prepared by chemically cross-linking IgG with DTSP to calibrate the sucrose density gradients which were used to separate IgGRF from IgA and IgM. In addition, the SDG fractionation was optimised using IgA, IgM and this mixture of IgG oligomers (Fig. 16,17) .


Preparation of RF IgG
RA serum
Pass through monomeric human IgG sepharose column
Wash column; elute RFs with sodum acetate pH3.5
Concentrate; apply to SDG pH3.5
Re-associate RF's at pH7 (glycine/HCl)
MBP/dot blot
Apply to SDG pH7
Analyse RF and RF complexes by SDS PAGE and RIA

Flow Chart 2

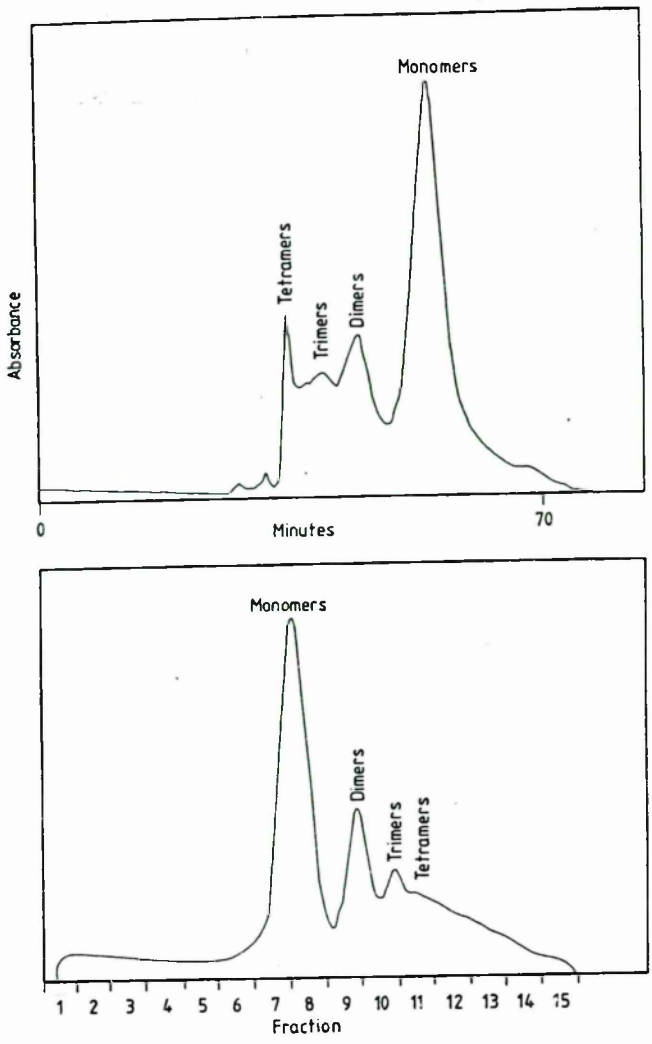


Figure 16: The fractionation of the chemically cross-linked oligomer preparation used to optimise and calibrate the SDG.

a. HPLC separation (TSK 3000) b. sucrose density gradient fractionation

4.3.2 Preparation of self associating IgG RFs (Flow Chart 2)

IgG RF in serum is complexed with IgG, IgA and IgM. IgGRFs either monomeric or self-associated or bound to other immunoglobulins were isolated by passing RA serum through a monomeric IgG sepharose affinity column. The unbound fractions were retained and from this the residual IgG was isolated (RF depleted IgG). IgGRF complexes were eluted from the column at pH 3.5; at low pH IgGRF monomers (7S) dissociated from IgG, IgA and IgM and were resolved by SDG (Fig. 17b,c). Sedimentation coefficients of the peaks were assigned by comparison both with the oligomer standards (Fig. 16) and with IgG and IgM fractionated under the same gradient conditions (Fig. 17a). IgG rheumatoid factor activity was confirmed by ELISA and the positive 7S IgGRF fractions were pooled. The 7S pool was adjusted to pH 7 and the sample again fractionated by SDG. This indicated that the RF monomers had self-associated to a mixture of monomers to tetramers (Fig. 17d,e).

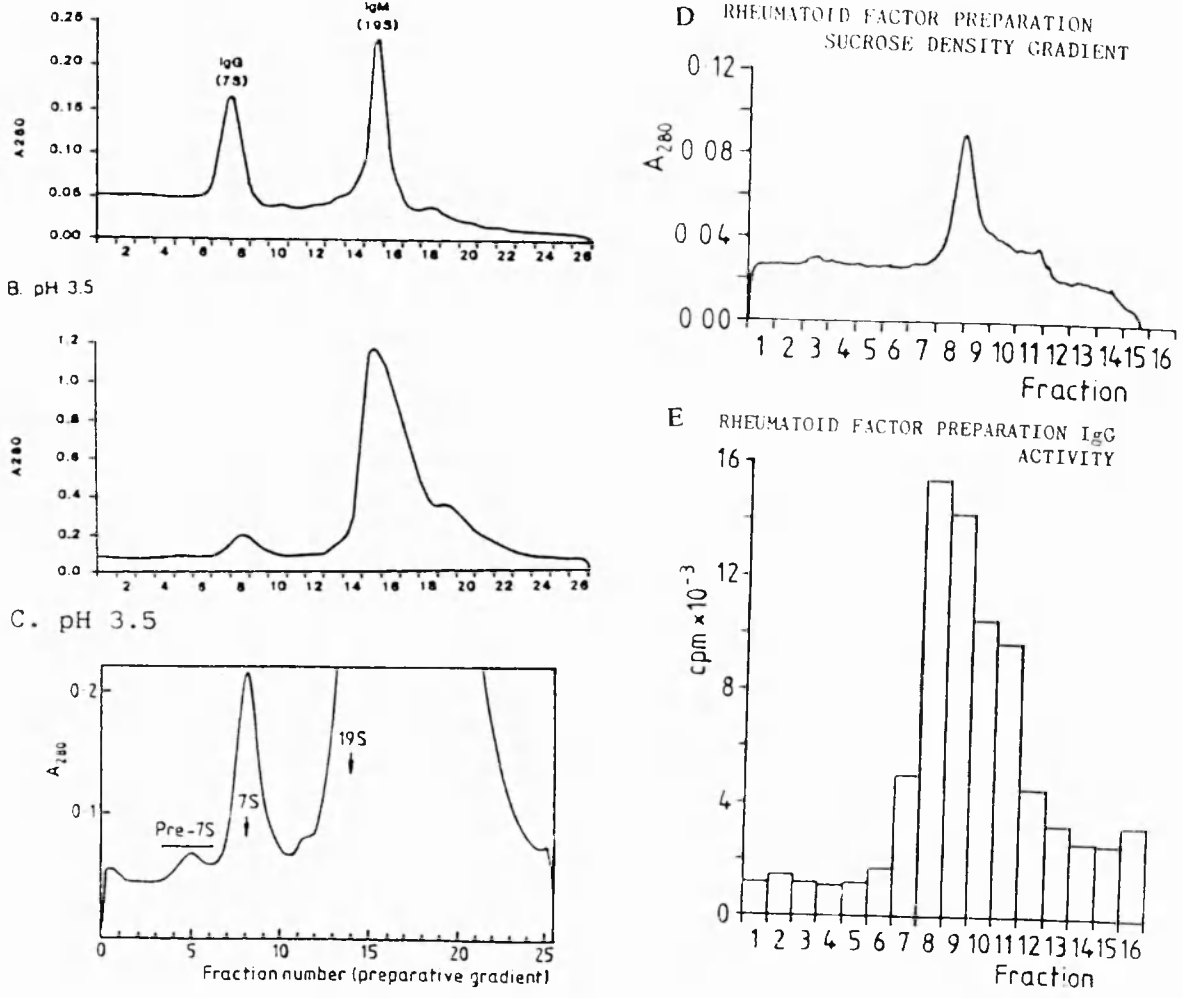


Figure 17: Fractionation of RFs by sucrose density gradient

(a) Standard markers: IgG (7s) and IgM (19s)

(b) RF preparation resolved at pH3.5

(c) Magnification of 1c to show resolution of <19s material. The 7S fraction contained IgGRF used in the experiments below

(d) pH7 SDG of re-associated aggregates

(e) Radioimmune assay of the fractions in 11d showing that the RFs have re-associated

4.3.3 Glycan analysis indicates that IgGRFs may both be glycosylated in the Fab region and contain increased levels of agalactosyl sugars

SDS PAGE of this preparation and others (Fig.9a) indicated that IgG RFs contained two heavy chains (Fig. 18) suggesting that up to 50% of IgG RFs are glycosylated in the Fab as well as the Fc, and that the proportion of extraglycosylated heavy chains increased with the size of the aggregates. This finding was confirmed and quantified by determining the number of sugar chains associated with Fab from RFs which was approximately 4 times as many as for normal IgG (Dr. S-C. Chang). Furthermore, glycan analysis of IgGRF showed that more than 85% of the sugars were of the G0 type (Dr. S-C. Chang). Together these data suggest that IgG RF consists of a specific set of IgG glycoforms which carry Fab glycosylation and have increased levels of agalactosyl sugars.

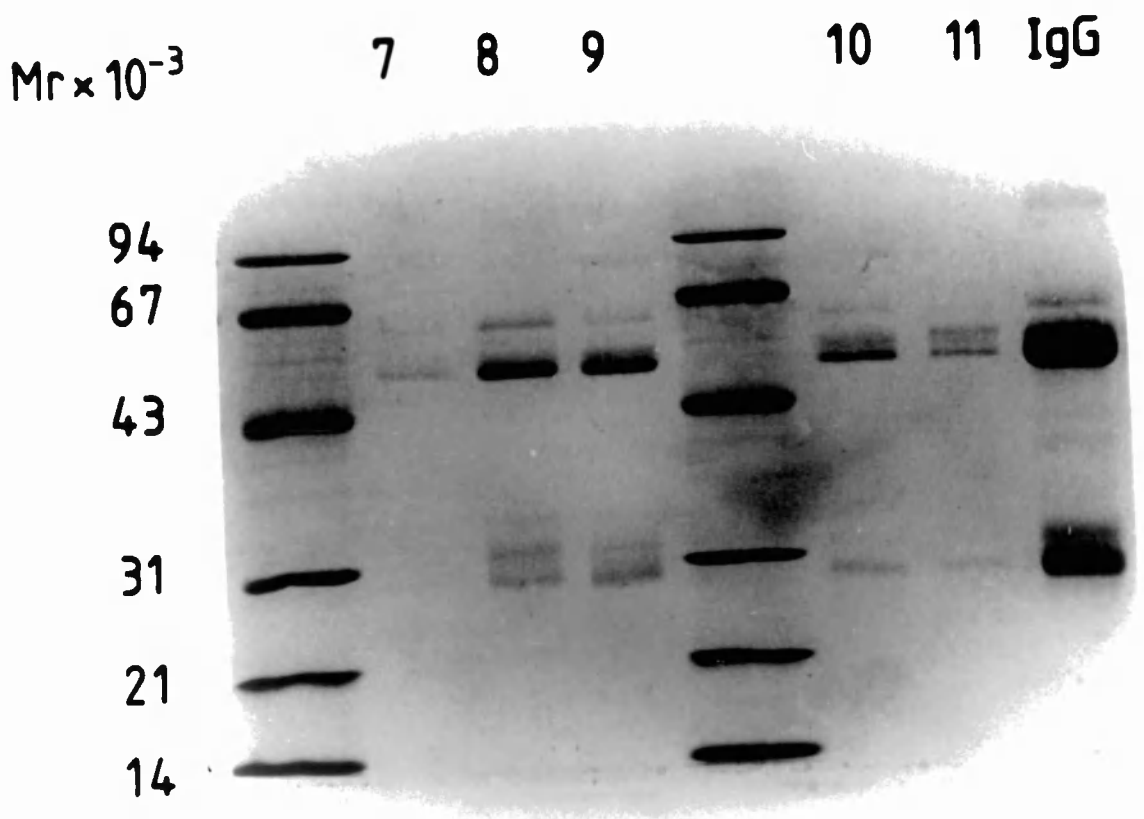


Figure 18: Analysis of RF aggregates

SDS PAGE of fractions from 11d showing that the proportion of extraglycosylated heavy chain increases with the size of the RF aggregates.

These data indicate that total IgG and IgGRFs from RA serum contain increased levels of IgG0, and that RFs have an increased proportion of Fab glycosylation. A model has been proposed to explain the formation of IgGRF complexes (Parekh et al 1985) which involves the recognition of galactosylated Fab sugars on one IgG molecule by a 'lectin-like' site available on the Fc of a second IgG molecule of the IgG0 type (See 1.2 and Fig.6b). The data in this section are consistent with this model, and moreover further studies (Scragg et al unpublished data) have shown that the increased level of IgG0 in RA IgG is mainly confined to the Fc region, and that the Fab region has a normal distribution of G2, G1 and G0 type sugars. This suggests that in RFIgG Fab sugars may have a role in recognition.

In the section which follows the possibility that the G0 oligosaccharides attached to IgGFc may be accessible for binding was probed with lectins, in the first instance as a means of isolating significant quantities of IgG0 glycoforms. Intact IgG and reduced and alkylated heavy chains from an RA patient were passed through an affinity column of bovine conglutinin coupled to CNBr sepharose. 70% of H-chains (48%G0) and 91% of the intact IgG (48%G0) was recovered in the flow through. None of the bound material could be recovered from the column by either

200mM GlcNAc, 10mM sodium barbital pH 7.4 or 0.1M sodium acetate. Conglutinin (which exists in bovines but not humans) binds terminal GlcNAc, mannose and fucose and has been shown to bind denatured IgG (Theil, S. personal communication), however data from these experiments suggested that although isolated heavy chains with G0 sugars may bind the lectin to some extent, the sugars on native IgG (48% IgG0) were not available for recognition under these conditions. However, the affinity of one CRD on conglutin for one ligand is very low (mmolar) and high affinity binding requires multiply presented sugars. In these experiments the IgG monomers in the fluid phase were not multiply presented which suggested that under these conditions the affinity of the IgG for bovine conglutinin may have been too low to allow significant binding.

4.4 Terminal GlcNAc residues on IgGRF and IgG0 may be exposed and available for binding

Another candidate for binding to exposed GlcNAc residues on IgG is a member of the collectin family, the serum mannose binding protein (MBP) which occurs in human serum. MBP is a Ca^{2+} dependent lectin with multiple CRDs each of which has a weak affinity for single terminal GlcNAc, mannose and glucose residues but which binds these sugars with high affinity if they are clustered together. In the next section human MBP was tested for binding to multiply presented IgGRFs containing >85% IgG0. Normal IgG, IgG RF and IgG prepared from serum depleted of IgGRF were immobilised to a blotting membrane to provide multiple presentation and tested for binding MBP. The physiological significance of this recognition event has not been investigated, however it may be noted that self associated IgGRF complexes are present in the synovial joints of RA patients, either as soluble aggregates or as deposits on the synovium (Leader et al - personal communication), and that MBP is present in RA synovial fluid (Malhotra et al 1995).

4.4.1 Preliminary experiments to investigate the affinity of normal IgG (G0 20%), IgG RF and IgG prepared from serum depleted of IgGRF (G0:45%) for serum MBP

5µl fractions of serial dilutions of normal IgG which contained from 10 to 0.001 µg/5µl and of IgGRF and IgG from RF depleted serum containing from 3 to 0.001 µg/µl were multiply presented by spotting onto a nitrocellulose blotting membrane. The membrane was exposed to serum MBP and the bound lectin detected with an polyclonal αMBP antibody. Normal IgG bound significantly at 10µg, and weakly at 3µg. IgGRF bound significantly at 3µg and at 1µg and very weakly at 0.3µg. IgG from serum depleted of IgGRF bound weakly at 3µg (Fig. 19). This suggests that IgGRF which contains self associating IgG and is therefore already multiply presented, binds MBP more efficiently than normal IgG which has been multiply presented only by concentration on the blotting membrane.

On page 172 line 3 should read:

'IgG glycoforms containing 100% IgG0 sugars were prepared by enzymatic modification of the natural....'

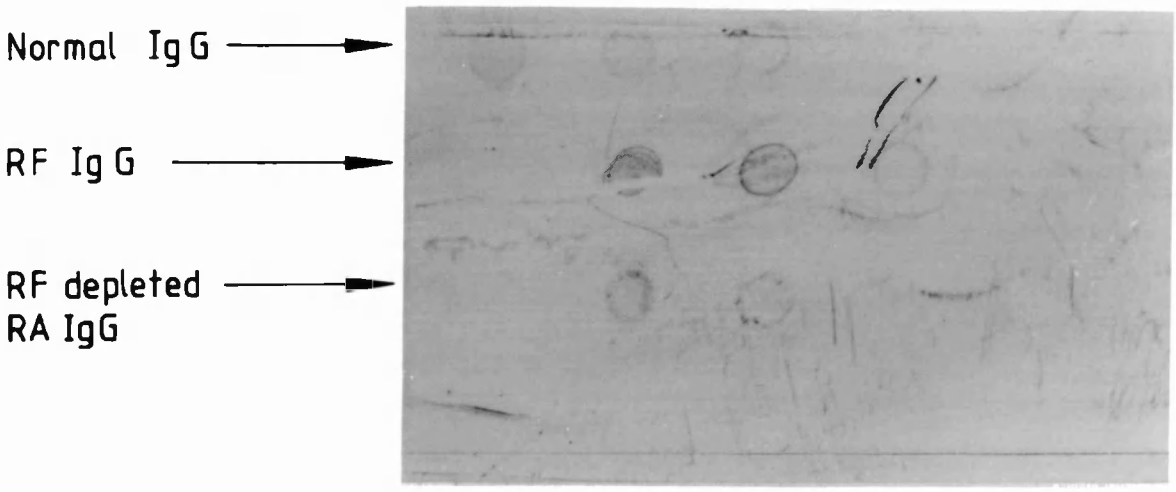


Figure 19: Dot blot showing that RFIgG binds MBP *at a 10-fold lower concentration than normal or RF depleted IgG. Top row contains a serial dilution of normal IgG; middle row contains IgGRF; bottom row contains the IgG remaining after the IgGRF had been removed. RF typically contains > 85% of G0 type glycans (S-C. Chang).*

As a result of these preliminary studies a more rigorous examination of the role of the sugars on IgG0 glycoforms in binding MBP was carried out in conjunction with Dr. R. Malhotra. A set of IgG glycoforms containing 100% G0 sugars by enzymatic modification of the natural population. 100% pure G0 glycoforms of Fab and Fc were also prepared to localise the binding site on IgG. The affinity of these specific sets of glycoforms containing only sugars terminating in GlcNAc was compared with the corresponding natural populations.

4.5 The preparation of IgG Fab and Fc by papain cleavage of IgG

There are three possible cleavage sites for papain in human IgG (Fig. 2b) one of which (224/5) is preferentially cleaved. The reaction conditions were optimised to reduce the cleavage at the more resistant sites to a minimum (none detected).

4.5.1 Papain cleavage of IgG

The optimised cleavage reaction, which separates IgG into Fab (M.Wt. 45-50kD) and Fc (M.Wt. 54kD), did not reach completion. This was found to be generally the case regardless of the cleavage conditions, and during the optimisation procedures the extent of cleavage was noted to be sample dependent (Fig. 20).

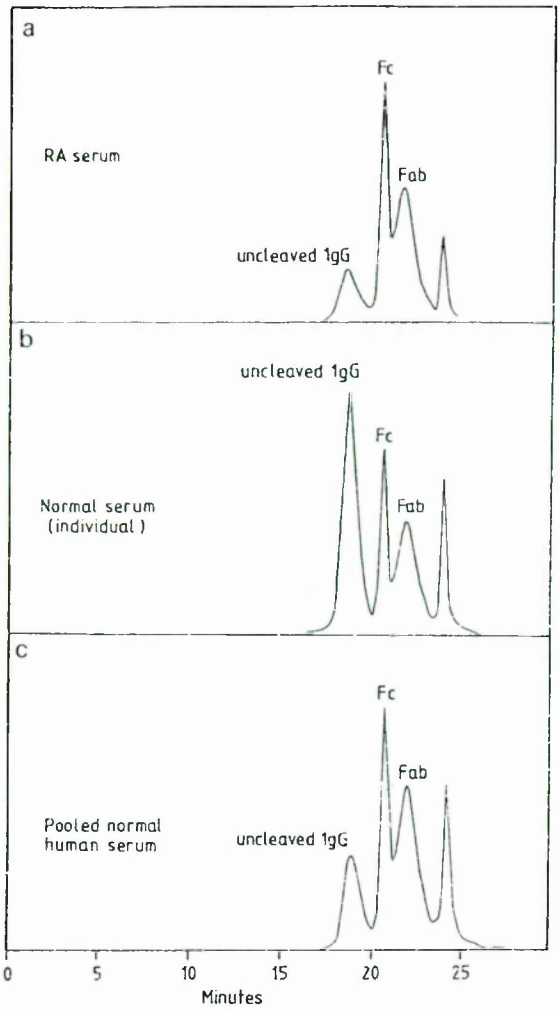


Figure 20: Products of papain cleavage of IgG from (a) an RA patient (b) a single normal individual and (c) a pool of normal serum *resolved by HPLC GPC (Tandem columns Zorbax GF 250 + 450; 2x9.4x250mm) showing the different susceptibilities of the three samples. In the normal IgG sample 18% of the total IgG population remains intact. In this particular RA IgG sample (G0 55%) 7.6% of the total IgG remained uncleaved.*

Papain cleavage rates are known to be subclass dependent, however it is not known whether there are variations in susceptibility between different glycoforms of IgG within the same subclass. In particular it is not known whether low levels of galactosylated sugars (high G0) in the Fc make the hinge region more susceptible to papain. Therefore prior to the preparation of Fab and Fc, the following study examined the relationship between the percentage of IgG0 sugars in 18 IgG samples and the percentage of uncleaved IgG.

4.5.2 IgG resistance to papain cleavage does not correlate with G0 levels

18 serum IgG samples were prepared from normal controls and patients with juvenile arthritis. The IgG samples, which contained a range of G0 values, were subjected to papain cleavage under identical conditions. The digestion products were resolved by HPLC GPC, and the percentage of uncleaved IgG plotted against the percentage of IgG0. (Fig. 21). These data suggest that there is no direct correlation between the two parameters.

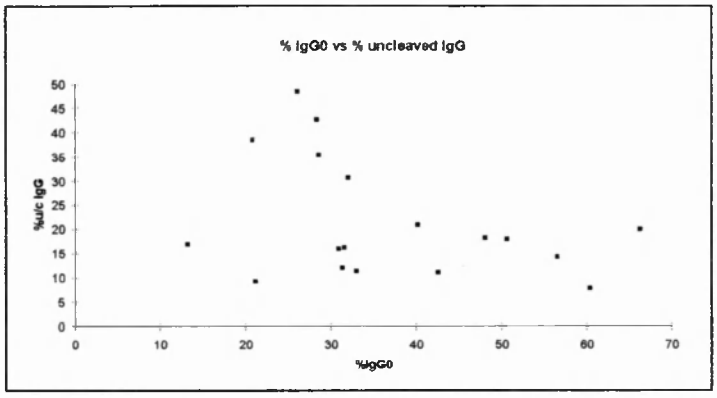


Figure 21: The % uncleaved IgG remaining after papain cleavage does not correlate with the G0 status of IgG. *The IgG samples were prepared from serum obtained from 14 patients with Juvenile RA and 4 normal controls. There was no significant correlation between the percentage of G0 sugars and the percentage of uncleaved IgG.*

4.5.3 Preparation and analysis of Fab and Fc from normal serum IgG (Flow chart 1)

Fab and Fc were prepared by papain cleavage of normal human serum IgG (35mg). Uncleaved IgG and Fc were separated from Fab by Protein G affinity chromatography. IgG and Fc were separated by gel permeation HPLC when any peptides resulting from the minor products of papain cleavage would also have been removed. The HPLC analysis of IgG is shown in figure 22a. There is some aggregated material in the sample which probably results from storage at -20°C (IgG is noted for the ability to form cryogenic precipitates). 25% of the IgG remained uncleaved. The HPLC analysis of the unbound fraction Fab (final yield 6mg; 34%) is shown in figure 22b. The non-selective losses mainly occurred as a result of conservative pooling to maintain purity and during concentration steps. The analysis of the bound fraction (uncleaved IgG+Fc) is shown in figure 22c. Fc (final yield 1.5mg; 18%) was separated from uncleaved IgG by HPLC GPC. The analysis of the purified Fc fragment is shown in figure 22d. Peaks were assigned with reference to known standards.

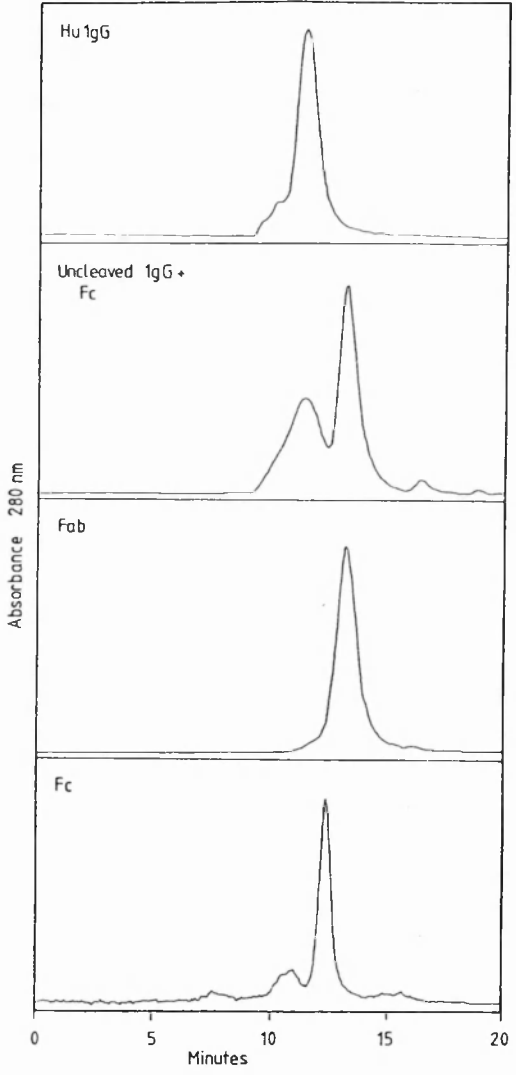


Figure 22: HPLC GPC (Zorbax GF250) analysis of (a) normal human serum IgG (b) purified Fab (c) uncleaved IgG and Fc eluted from a protein G affinity column (d) purified Fc. *The variable region of the Fab has a range of molecular weights due both to variations in the polypeptide and the possibility that the region may be glycosylated. This results in a broad elution band both for Fab and for intact IgG. In contrast, the Fc region is highly conserved and the elution profile correspondingly narrow.*

4.5.4 Charge analysis of normal IgG, Fab and Fc sugars

The glycans were released from 200mg of normal IgG, Fab and Fc by hydrazine. The pool of glycans was labelled with 2AB and analysed by WAX (Fig. 23a-c) according to the procedures used earlier in this chapter. Comparable fluorescence for the Fab, Fc and IgG glycans was achieved by injecting a fraction of the Fab sugars corresponding to 5x as much protein as was required for IgG and Fc. This is consistent with the fact that only 20% of IgG molecules are glycosylated in the Fab region. Peaks were assigned by comparison with standard fetuin sugars resolved under the same conditions (Fig. 23d). The relative proportions of neutral:A1:A2 glycans is shown in table 2. Fab contains a higher proportion of charged glycans, in addition most of the disiaylated glycans are associated with Fab.

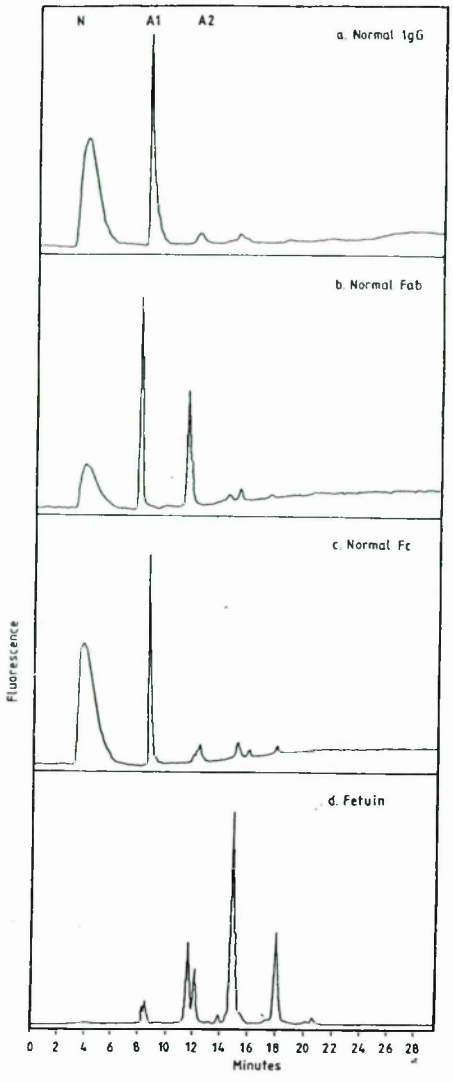
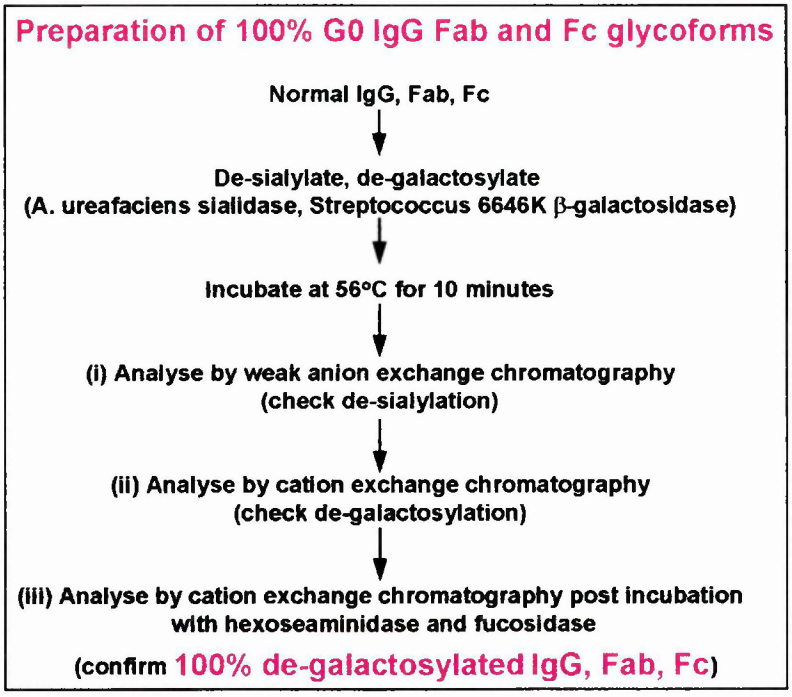


Figure 23: a-d sugars released from normal IgG (a), Fab (b), Fc (c) and fetuin (d) were resolved by WAX chromatography. After treatment with *Arthrobacter ureafaciens* sialidase all structures became neutral indicating that all charge is due to sialic acid.

	neutral	%A1	%A2
Normal IgG	65.3%	28.3%	6.4%
Fab	40%	30%	30%
Fc	68%	27.4%	4.6%

Table 2: The percentages of neutral, mono (A1) and di-sialylated (A2) glycans attached to normal IgG, Fab and Fc.


Preparation of 100% G0 IgG Fab and Fc glycoforms
Normal IgG, Fab, Fc
De-sialylate, de-galactosylate
(A. ureafaciens sialidase, Streptococcus 6646K β-galactosidase)
Incubate at 56°C for 10 minutes
(i) Analyse by weak anion exchange chromatography
(check de-sialylation)
(ii) Analyse by cation exchange chromatography
(check de-galactosylation)
(iii) Analyse by cation exchange chromatography post incubation
with hexoseaminidase and fucosidase
(confirm 100% de-galactosylated IgG, Fab, Fc)

Flow Chart 3

4.6 Preparation and analysis of IgG0, FabG0 and FcG0 glycoforms *(Flow Chart 3)*

4.6.1 Preparation of IgG0, FabG0 and FcG0

Normal serum IgG (IgG0:20%) (400μg) and purified Fab (800μg) and Fc (200μg) were de-sialylated and de-galactosylated by incubating the proteins with *Arthrobacter ureafaciens* neuraminidase and *Streptococcus* 6646k. Concentrations were kept low (0.5mg/ml) to avoid aggregation of the G0 glycoforms which was noted in other studies. Oligosaccharides were analysed for G0 by cation exchange chromatography (Fig. 24) following hydrazine release, labelling and enzymatic removal of terminal GlcNAc and fucose. In each case the oligosaccharides associated with the molecules were 100% G0 type. A number of controls were carried out to confirm the activity of the enzymes:

(i) Prior to the digestions the activity of *Streptococcal* 6646K β-galactosidase was tested with free asialo G2 sugars. 100% were reduced to G0.

(ii) Sialidase activity was checked by analysing the sugars released from IgG, Fab and Fc treated with *Arthrobacter ureafaciens* sialidase and *Streptococcal* 6646K β-galactosidase by WAX chromatography. 100% of the structures were neutral.

(iii) The sugars released from IgG, Fab and Fc treated with *Arthrobacter ureafaciens* sialidase and *Streptococcal* 6646K β-galactosidase were analysed by cation exchange chromatography (Fig. 24) and MALDI MS to confirm galactosidase digestion was complete.

(iv) 100% de-galactosylation was confirmed by treating the released glycans with the G0 standard mix to remove core fucose and bisecting GlcNAc residues. All populations eluted as a single peak on the cation exchange column.

4.6.2 Analysis of 2AB labelled G0 glycoforms of (a) IgG (b) Fab (c) Fc by cation exchange chromatography

Figure 24 shows the analysis of the sugars released from IgG, Fab and Fc after treatment with *Arthrobacter ureafaciens* sialidase and *Streptococcal* 6646K β-galactosidase. 90% of IgG and Fc sugars elute as a single peak at the elution position of standard G0. Over 90% of the Fab sugars, which contained two major and two minor populations in figure 13c, eluted as a single peak (d) after treatment with the standard G0 mix (ch.8 section 1.4h) confirming that the sugars were all of the G0 type. 80.6% of IgG sugars are fucosylated and 24.5% are bisected (Table 1b). 70% of Fab structures are bisected, (Parekh) and Fab contains relatively few fucosylated structures. This suggests that the major peaks in 13c may represent the G0 biantennary complex glycan +/- bisecting GlcNAc-fucose (+B-F; -B-F). The major peak in IgG and Fc is the core fucosylated G0 biantennary complex glycan (-B+F). Of the total IgG oligosaccharides, on average Fab sugars account for only 20% of the total glycans, the remaining 80% are derived from the Fc. The relative proportion of Fab:Fc sugars were estimated from these data and were consistent with this finding. The sugars from each were released from equal amounts of protein. As in 5.3.4 it was necessary to inject sugars from 5x as much Fab protein as for Fc and IgG (125μl : 25μl : 25μl) to obtain the same amount of absorbance, indicating that

approximately 20% of the Fab chains in this sample were glycosylated. Subsequently the natural mixture of IgG, Fc and Fab glycoforms and the pure set of IgG, Fc and FabG0 glycoforms together with RA IgG were tested for their affinity for MBP and to activate complement. The analysis examined the possiblity that the N-acetyl glucosamine residues which terminate all the oligosaccharides present in the set of IgG0 glycoforms may have a role in recognition.

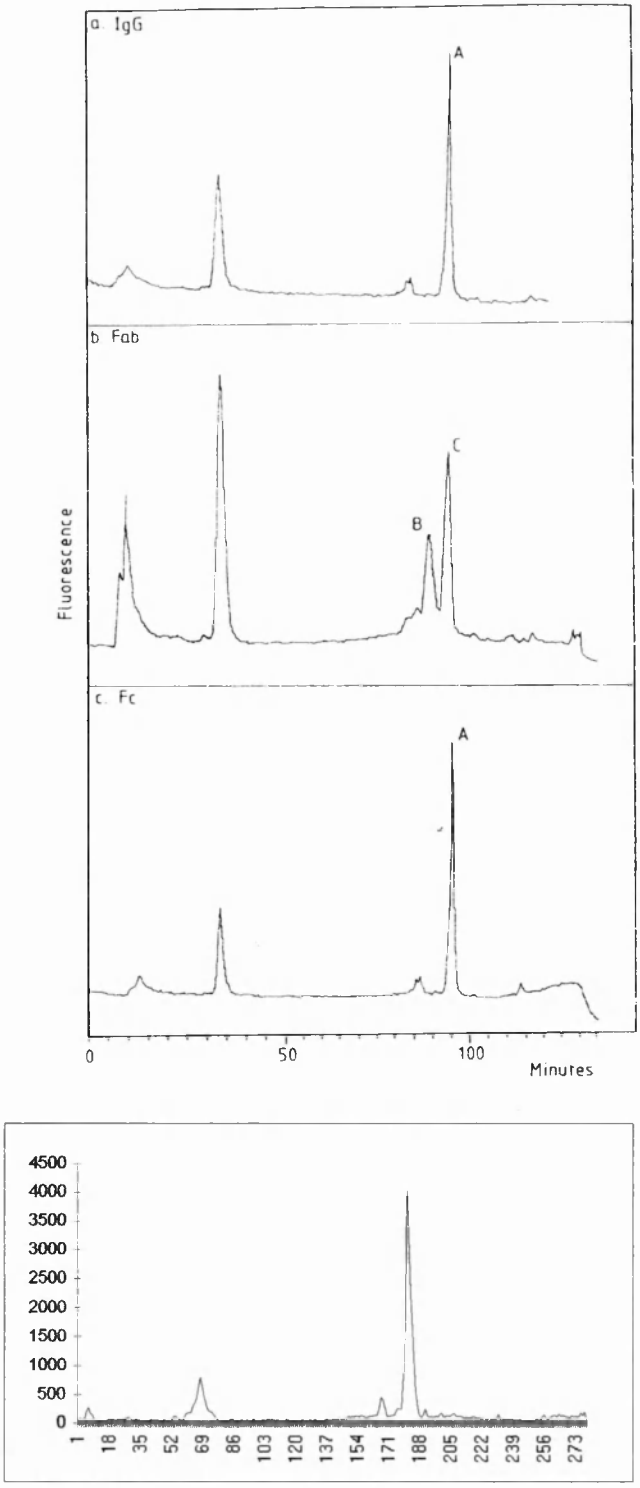


Figure 24: Cation exchange chromatography of 2AB labelled sugars released from (a) IgG0 (b) Fc G0 (c) Fab G0 and (d) tritium labelled Fab sugars post G0 mix. *The*

assignment of the peaks is discussed above; by comparison with known standards and the profile in figure 15a and the MS analysis of the IgG glycans in figure 12 all are assigned to the G0 type. In (a) and (c) peak A is the complex biantennary glycan with a core structure (-B+F). In (b) the earlier peak, B, contains the complex biantennary glycan with a core structure (-B-F) and the later peak, C, contains the complex biantennary glycan with a core structure (+B-F). (d) confirms that after incubation of (b) with the G0 mix (containing fucosidase and hexosaminidase) the two glycans are reduced to a single structure eluting at the position of G0.

4.7 The interaction of MBP with IgG, IgG0 and their fragments and IgG induced activation of the complement system

To investigate systematically the possibility that MBP might bind IgG0 and thus activate the complement pathway, and to localise the binding site for MBP on IgG, the following compounds were tested both in MBP affinity and in C4b deposition assays: (i) Normal IgG (20% G0) (ii) RA IgG (36%G0) (iii) natural glycoforms of normal Fab and Fc (iv) the 100% G0 subsets of glycoforms, IgG0, FabG0 and FcG0. The assays were performed with IgG bound to microtitre plates and therefore examined the binding of multiply presented molecules. The mannose binding assay determined the relative amounts of MBP bound to microtitre plate coated with each of the arrays of glycoforms of IgG, Fab in (i-iv) above using a monoclonal antibody to MBP. In the C4b deposition assays microtitre plates were coated with IgG, IgG0 or their fragments (i-iv) and vacant sites blocked with BSA. Mannose binding protein was incubated in the wells and the plates were then washed. MBP/C1q depleted serum was then incubated in the coated wells. The assay tested whether MBP bound to exposed N-acetyl glucosamine residues on IgG, Fab or Fc on the plates could activate the complement pathway (Fig. 27) by assaying for C4b deposits. Like C1q, MBP can activate the pro enzymes C1r and C1s. It can also act via the mannose binding protein associated serum protease, MASP, which appears to substitute for C1r and C1s (Matsushita and Fujita 1992). C1s cleaves C4 to form C4a and C4b. C4b undergoes a conformational re-arrangement which exposes a thioester group (-S-C(=O)-C) which is highly susceptible to nucleophilic attack. The electron deficient C in the thioester bond is highly reactive and readily forms a covalent bond with electron donating groups, in particular OH or NH_2. If C4b is generated in the assay, much of it (approximately 90%) will be hydrolysed by water, but some will bind to hydroxyl or amino groups which are in close proximity (<400A) in the protein bound to the microtitre plates. The deposited C4b was then detected with a monoclonal antibody.

The data were obtained by Dr. R. Malhotra. The main conclusions were:

(i) The calcium dependent binding of MBP to IgG0 or IgG was inhibited in a concentration dependent manner by mannose.

(ii) In the presence of Ca^{2+} MBP binds IgG0 (G0=100%) > IgG from an RA patient (G0=36%)> normal IgG (G0=20%).

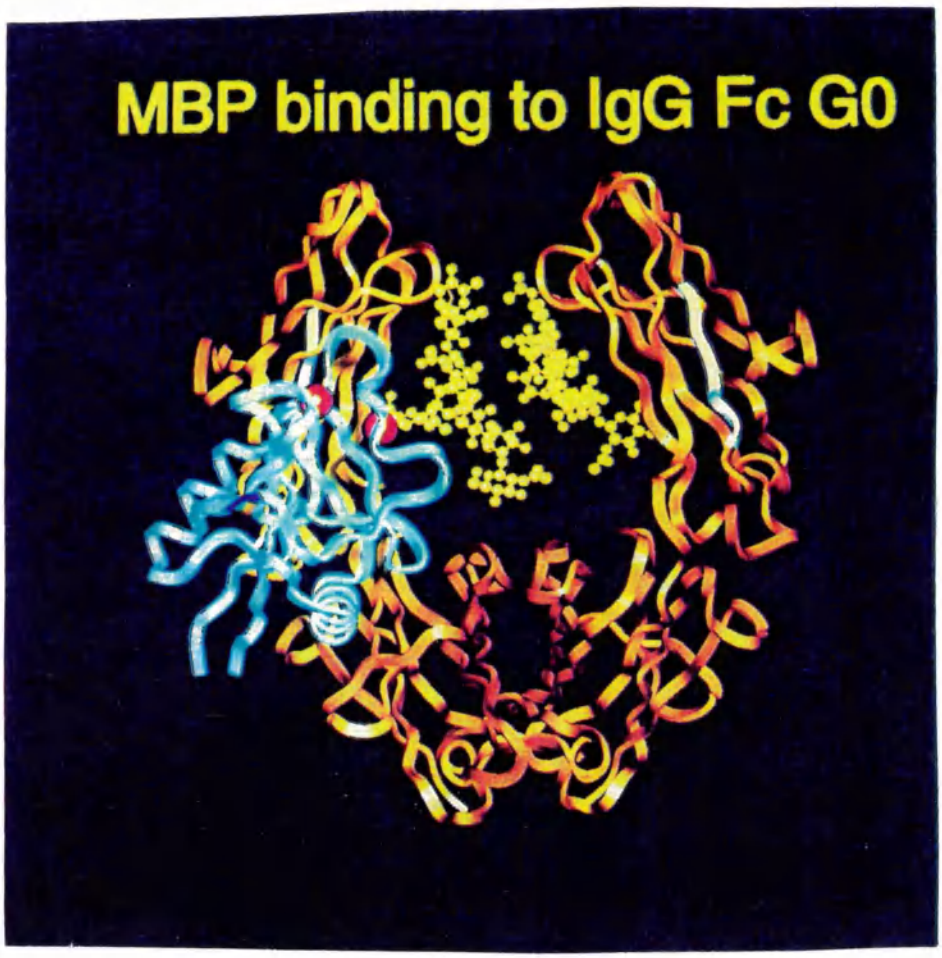


Figure 25: Molecular model of FcG0 and a single carbohydrate recognition domain (CRD) from MBP *The model is based on information from the MBP crystal structure (Weiss et al 1992) and NMR studies of the binding of free monosaccharides to the CRD (lobst et al 1994). The Fc structure is based on the crystal structure of Deisenhofer (1981) and the CRD is based on the crystal structure (Weiss et al 1992). The α1,6 non-reducing terminal GlcNAc residue has been docked to the CRD saccharide binding site. The principle interaction involves the chelation of the two cis 3-4- hydroxyl groups of the GlcNAc residue from the Fc to the Ca^{2+} ion of the CRD. The CRD can also interact with the 3-arm non-reducing terminal GlcNAc residue by an alternative displacement of the Fc oligosaccharide chain. The C1q binding site on the right hand side of the Fc CH_2 domain (Burton et al 1980, Duncan and Winter 1988) is also shown. (Malhotra et al 1995)*

(iii) G0 glycoforms of normal IgG and Fc, but not Fab, show a higher Ca^{2+} dependent affinity for MBP than the unmodified normal populations of IgG and Fc and Fab.

(iv) IgG0 induces 5x more deposition of C4b than IgG

(v) Increase in C4b deposition occurs when Fc is converted to FcG0. There is no change in C4b deposition occurs when Fab is converted to FabG0

(vi) Mannose significantly decreased the deposition of C4b suggesting that binding of MBP is through the oligosaccharide.

4.8 Molecular modelling of FcG0 and a single carbohydrate recognition domain (CRD) from MBP (In conjunction with Dr. Mark Wormald)

The model (Fig. 25) indicates that, in IgG0 glycoforms, terminal GlcNAc residues in the Fc are exposed and available to interact with the MBP lectin. This suggests a molecular basis for the experimental data which is based on NMR studies of the solvent accessible surface (Malhotra et al 1995). These indicate that in the G0 glycoform the glycans are released from the protein surface, and therefore in the model the G0 oligosaccharide has been moved freely relative to the protein around the Asn297 C-C bonds.

5. Discussion

5.1 Novel sequencing strategies

The development of automated methods for releasing and profiling oligosaccharides has highlighted a number of other areas in which conventional sequencing technology could be improved. During the work described in this chapter alternative methods for labelling, profiling and sequencing were explored and used, in conjunction with conventional sequencing, to analyse the glycans associated with IgG.

5.1.1 Fluorescent labelling of the reducing terminus of the glycans coupled with on line detection reduces artefacts, improves data analysis and reveals more information.

Conventionally tritium is used to label the released glycans at the reducing terminus. Reduction of released glycan pools with ^3H_4B is non-selective, and the monitoring systems allow nanomoles of sugar to be detected. However, peptide fragments co-purifying with the oligosaccharides may inadvertently be labelled, giving rise to artefacts and causing problems with interpretation, particularly in charge analysis. The use of the fluorescent compound 2-aminobenzamide, which also labels oligosaccharides non-selectively (Hase et al 1976), has decreased the number of artefacts and extended the limits of detection by about 10-fold. In this laboratory, fluorescent labelling has allowed on line detectors to be coupled to HPLC columns eliminating the need for hand counting fractions. A comparison of the data in figure 10a,b (radiolabelling and hand counting) with 10c,d (fluorescent labelling and on line detection) shows that the on-line detection system, in which the fluorescent eluate is monitored 20 times a second, records the full peak heights which have been missed in the hand counted 15 sec radiolabelled fractions. The improved data analysis has important implications. In this case it

revealed that the A1 glycans associated with RA IgG consist of a mixture of G1 sugars in which the single terminal galactose is sialylated and G2 sugars in which only one of the terminal galactose residues has been sialylated. In contrast, the profile for normal IgG glycans indicates that the normal IgG A1 glycan pool contains only monosialylated G1 structures.

5.1.2 MS analysis eliminates the need to label and separate glycans.

The need for labelling has been eliminated altogether by the use of MALDI MS (Fig.12) which gives molecular weights to approximately +/-0.3 of a mass unit. The released glycans were desialylated, passed through mixed bed ion exchange resins and analysed directly by MALDI MS. Although the glycans may still contain peptide fragments as a result of the hydrazinolysis procedures, the relatively large and anomalous molecular masses of such artefacts make it improbable that they will be confused with sugars. The use of MS in this thesis is discussed in chapter 8 section 2.1.

5.1.3 WAX chromatography replaces HVE for charge analysis of the released glycans.

In conventional sequencing charge analysis involved separation of the glycans by high voltage electrophoresis (HVE). In this thesis, this technique has been replaced entirely by WAX HPLC which is preparative, fast, highly reproducible and gives improved resolution.

5.1.4 Cation exchange (CAT) chromatography has been used to analyse asialo glycans.

Figure 15a,b shows the quality of the resolution which has been achieved by HPLC (Vydac 400VHP575) CAT chromatography. The technique is preparative, fast and highly reproducible over long periods of time and from column to column. Currently peaks are assigned by comparison with known standards; in addition to hydrodynamic volume other properties of the glycans, such as branching, influence the elution time. Although this makes it more difficult to predict elution times and to relate them to a fundamental property, CAT HPLC provides a useful alternative strategy to P4 gel filtration, which can at times give poor resolution (see Fig. 11a,b). The soft P4 gel is technically difficult to handle. For example, it needs to be fined to select the narrow size range of particles which will resolve glucose oligomers of up to 24 glucose units. In addition, the columns need to be maintained at constant temperature (50°C) and pressure and resolution is lost if the gel collapses, and this can happen even in newly poured columns.

5.1.5 Simultaneous exoglycosidase digestion of mixtures of glycans using MALDI MS to analyse the digests.

This novel approach to sequencing was developed to address the difficulties encountered in conventional exoglycosidase sequencing strategies which require the preliminary separation of the glycan pool into pure components using P4 GPC. This may be technically difficult because the individual components often elute in overlapping peaks. This necessitates selective pooling of central peak fractions and re-chromatography, with the result that much of the sample is lost.

Once purified, sequencing individual glycans requires multiple manipulations, even with the advantages of the RAAM technology described in ch.8 section 1.8. In this chapter (figure 12) MALDI MS has been used to determine the molecular weights, and therefore the composition, of the components of the glycan pool in a single analysis and subsequently to follow the sequential exoglycosidase digestions of each oligosaccharide simultaneously. Although arm specificity cannot be determined from this strategy, linkage information is obtained from the use of specific exoglycosidase enzymes. A combination of conventional and novel sequencing strategies was used to analyse the glycans associated with IgG.

5.2 Site specific glycosylation of IgG, Fab and Fc and IgGRF

5.2.1 The normal human asialo IgG glycan population is composed of a mixture of 16 complex biantennary glycans

16 glycans were identified in the mass spectrometry profile of the unlabelled asialo glycans released from normal IgG (Fig. 12, Table 1). One of these, H3N3 is the unfucosylated form of a structure previously noted by Jefferis (H3N3F), and has been previously reported (Harada et al 1987). Other studies of the glycosylation of human IgG include those carried out by Mizouchi et al 1982, in which truncated, hybrid and mannose structures were detected at very low levels. In the analysis in this chapter some of these structures are present at low levels (<2%). Although they have also been reported by Jefferis, the possibility that, in this preparation, they are derived from low level contamination of the IgG cannot be ruled out. Some details of the analysis of polyclonal human IgG analysed by Jefferis (1990) and Parekh (1988) using HPLC and P4 GPC respectively are given in Table 1. Table 3 compares the relative proportions of the different classes of sugars identified in this chapter and in the two earlier studies. These tables suggest that the percentages of IgG0, G1 and G2 glycoforms remain relatively constant from sample to sample. However, it is not valid to draw such conclusions from an analysis of three samples. In a study of 400 normal controls carried out in this laboratory (Parekh et al 1988a) the percentage of IgG0 was shown to be an age related parameter. The percentage of IgG0 in normal IgG is lowest in the mid twenties (typically 20%) and rises with age to around 35-40%. IgG2 varies in inverse proportion to IgG0 while IgG1 remains relatively constant.

	G2	G1	G0	Fucosylated	Bisected
(i) MALDI MS	41.7%	38.7%	19.6 %	80.6%	24.5%
(ii) G0 mix/CAT	41	39	20		
(iii) HPLC (Jefferis)	37.8	39.9	22.3	81.9%	21.5%
(iv) P4 GPC (Parekh)	37.9	37.2	23.7		

Table 3: The classes of sugars on normal IgG determined by (i) MALDI MS (ii) cation exchange chromatography after digestion of fucose and GlcNAc using the 'G0 mix', (iii) HPLC (Jefferis et al 1990) (iv) P4 GPC and exoglycosidase sequencing (Parekh 1986)

5.2.2 Site specific glycosylation of IgG may partly be determined by the local protein structure

The local 3-dimensional structure of the protein may be a factor which controls the levels of galactosylation and sialylation of Fab and Fc glycans. 76% of the glycans released from normal IgG were neutral, 12% were monosialylated and 10% were disialylated (Fig. 10 a,b). 95% of the disiaylated glycans were in the Fab fragment (Fig 23 a-c). This site specific glycosylation suggests that the protein structure has a role in directing its own glycosylation by restricting access to sialyl transferases to the developing glycan chain in the Fc. The increase in Fab sialylation over Fc reflects the higher percentage of galactose in Fab sugars, and this also suggests that there is restricted access of enzymes to the Fc glycans during processing. In the ER IgG is assembled completely with two heavy and two light chains prior to transfer to the Golgi apparatus where complex processing occurs (Bole et al 1986). Thus when the glycoprotein is exposed to galactosyl and sialyl transferases the Fc sugars are already confined within the interstitial region between the two CH2 domains where they interact with the protein surface. The terminal residues are therefore protected from glycosyl transferases by the 3-dimensional structure of the CH2 domains and their flexibility is restricted by oligosaccharide-protein interactions. The limited space does not significantly reduce the affinity of the N-acetylglucosaminyl transferases I and II, although a small percentage (2-3%) of normal IgG molecules have been detected without substitution of GlcNAc residues on the 1,3 arm of the trimannosyl core. However as the chain elongates the interstitial space decreases and the number of protein oligosaccharide interactions increases and the glycosylating enzymes become increasingly inefficient. As a result a significant proportion of normal Fc sugars are not fully galactosylated and the majority lack sialic acid. To some extent the sialylation of the Fab sugars is also restricted by the local three-dimensional protein structure; free light chains in BenceJones protein are heavily sialylated (>85%) indicating that in IgG the three-dimensional structure of the Fab region may influence the activity of the sialyl transferase. Serum glycoproteins are generally heavily sialylated, and terminal sialic acid is generally understood to prevent recognition and clearance by the asialo glycoprotein receptor. This implies that the local protein structure also has a role in reducing the affinity of the sugars on IgG for this receptor.

Another feature of site specific glycosylation in IgG involves the distribution of the types of core structures. In Fc the complex biantennary glycan which predominates has a single core type (-B+F) which is fucosylated (F) but does not contain a bisecting GlcNAc (B). In Fab two major core structures were detected: the data suggested that these are -B-F (approximately 40%) and +B-F (approximately 60%). The means by which the cell and the protein structure control the site specific glycosylation of the glycan cores are not known, neither are the consequences.

5.2.3 B-cell transferases control the levels of IgG0 in normal and RA IgG

In addition to the constraints imposed by the structure of the individual protein, the cell plays a major role in defining glycosylation by controlling the levels and types of glycosylating enzymes. B-cells are a mixture of clones which have different sets and ratios of glycosyl

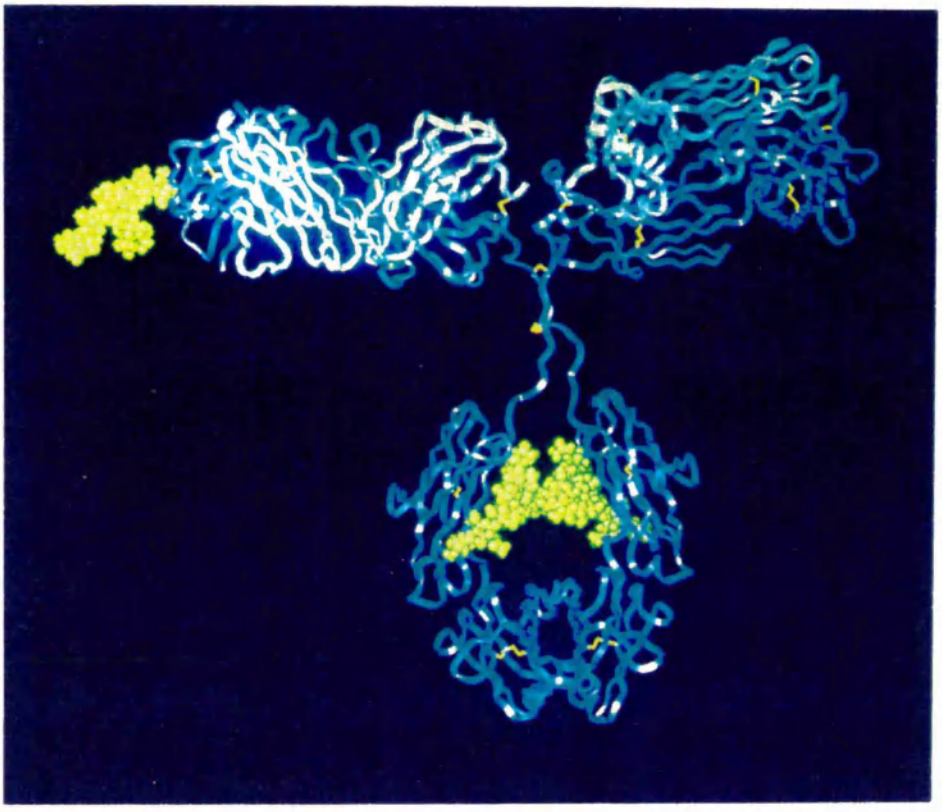


Figure 26: A model of IgG subclass1 showing Fab glycosylation

In addition to the conserved glycosylation site at Asn297 in the CH2 region of the heavy chain some IgG molecules contain glycosylation sites in the variable region of the Fab. This gives rise to heavy chians containing more than one oligosaccharide. Extraglycosylated heavy chains are associated with IgGRF complexes (Fig.18) and it has been proposed that such glycoforms may be involved in complex formation (Parekh et al 1984).

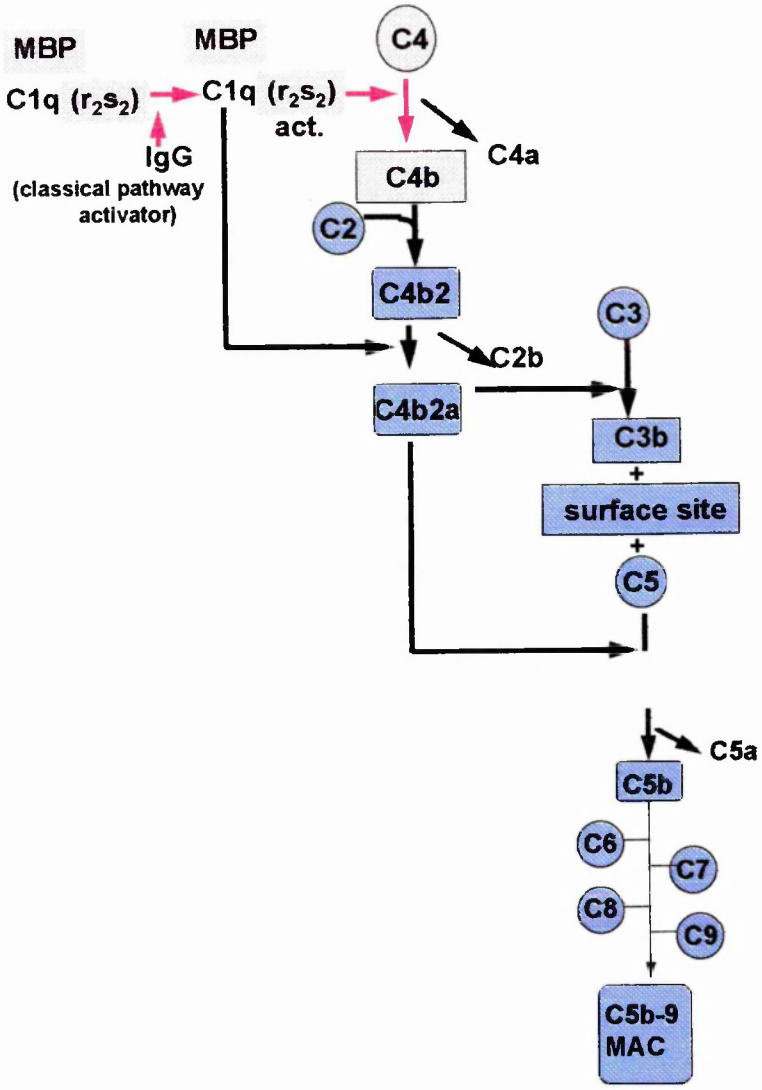
transferases. The difference in levels of IgG0 glycoforms of IgG in normal (20%) and RA (36%) noted in this chapter (Fig.15a,b) and elsewhere are not reflected in comparisons of other normal and RA serum glycoproteins such as transferrin (Endo and Kobata - unpublished data). The increased levels of G0 in RA IgG are therefore more likely to result from low levels of galactosyl transferase (Gal-T) activity than from de-galactosylation of IgG in the circulation. This has been demonstrated by Furukawa et al (1989): the low activity levels are due to the fact that there is a specific GalT enzyme for IgG and that the apparent km of GTase in RA IgG B-cells for UDPGal is three times the level in B cells from normal serum. It has been suggested that a reason for this low activity may be that RFs in RA B-cell homogenates may bind to acceptor asialoagalacto IgG and inhibit transferase activity. However microsomal proteins (containing no RF) extracted from RA B-cells were also unable to galactosylate IgG as efficiently as the enzyme from normal B-cells. This result suggests that the cells producing IgG with high G0 are a specific subset of B-cells and that IgGRFs, discussed below, are also the product of B-cell specific clones.

5.2.4 IgG RFs contain increased levels of Fab sugars and G0 compared with normal IgG

IgGRF contains two species of heavy chain (Fig.18). Both bands were reduced to a single band of approximately 47kD following digestion with PNGase F (data not shown) indicating that the heavier of the two chains is extraglycosylated in the Fab region. The relative proportion of this band increased to 50% of the heavy chain population in larger RF complexes, suggesting that Fab glycosylation (Fig. 26) may be involved in the formation of self associated IgGRF complexes. IgGRF monomers and complexes are present in the serum and synovial fluid of RA patients and a study by Chang (1993) indicated that Fab isolated from RA serum IgG contained more than four times as many sugar chains as normal Fab. IgGRF also contains increased levels of G0 sugars (>85%; Chang 1993). Together these data suggest that RF complexes are composed of a subpopulation of IgG molecules which are glycosylated in the Fab region and have a high percentage of G0 sugars in the Fc. A proposed model for IgG RF aggregates (Roitt et al 1988) (Fig 6b) involves the interaction of galactosylated Fab sugars from one IgG molecule with the CH2 domain of a second. This requires a site to be exposed on the CH2 domain which is able to participate in intermolecular oligosaccharide-protein interactions; such a site may be available in agalactosyl IgG. While an increased level of IgG0 in serum constitutes a risk factor for the development of RA (Tomana et al 1994) and is an indicator of disease severity (Parekh et al 1988b) the possibility of a role for IgG0 and IgG RF in the pathogenesis of the disease continues to be explored.

5.3 IgG0 sugars can be recognised by mannose binding protein (MBP)

A route for activation of the classical complement pathway (Fig. 27) has been identified which involves the Ca^{2+} dependent lectin, MBP. MBP contains multiple polypeptides which are organised into subunits containing triple helical stalks in the collagen - like region of the molecule and globular heads in the C-terminal domain (Reviews: Weiss and Drickamer 1994,



Activation of the classical complement pathway by C1q or MBP

Formation of C5-9 MAC complex which binds CD59 (ch2,3)

Figure 27: *IgG0 was bound to a microtitre plate and MBP incubated in the wells. Serum depleted of MBP and C1q was then incubated in the coated wells and the extent of complement activation assessed by assaying for C4b deposits (Malhotra et al 1995). The mauve colour indictes those molecules involved in the activation of complement by IgG0 binding MBP.*

Hoppe and Reid 1994). Each subunit carries two carbohydrate recognition domains, and one molecule of serum MBP contains up to 18 identical CRDs. The affinity of one CRD for a single terminal monosaccharide is relatively weak (kd:1mM) and to activate complement MBP requires multiple ligands to bind to the clustered CRDs. The large distance between the binding sites (5.3nM) (Weiss and Drickamer 1994) makes it impossible for a single mammalian high mannose oligosaccharide to interact with more than one CRD within the trimer. In serum the multiple binding of oligomannose cell surface oligosaccharides on various bacteria and fungi by MBP leads to their neutralisation by complement - mediated cell lysis or opsonisation. MBP binds terminal N-acetyl glucosamine, mannose, glucose and fucose residues. In this chapter IgG0, FcG0 and FabG0 glycoforms, which contained only oligosaccharides terminating in GlcNAc, were prepared. These pure glycoforms were used in binding studies which showed that multiply presented IgG0 and FcG0 bound MBP and fixed C4 more efficiently than multiply presented normal IgG, Fab, Fc or FabG0. This suggested that the G0 oligosaccharides on Fc were involved in the recognition event. Molecular modelling studies (Fig.25) indicated that terminal GlcNAc residues on either the α1,3 or α1,6 arms of the Fc oligosaccharides are exposed and available for binding the MBP. In normal human IgG 80% of galactose in G1 type sugars is associated with the α1,6 arm. It is not known whether this ratio is altered in RA IgG. Interestingly, FabG0, which also contains exposed GlcNAc residues, did not interact significantly with MBP. The affinity of a single CRD for a single sugar residue is weak (1 mmolar), therefore to achieve high affinity binding multiple CRDs on MBP must interact with multiply presented ligands. In the *in vitro* studies IgG, IgG0 and the corresponding Fab and Fc fragments were multiply presented on the surface of the microtitre plate. Only 20% of Fab sugars are glycosylated therefore the density of terminal GlcNAc residues presented by FabG0 was significantly less than for FcG0 and IgG0 in which every molecule contains two (or more in the case of IgG) G0 sugars. This suggests that, if these data are to be relevant *in vivo*, multiple presentation of the G0 sugars attached to many IgG molecules may be necessary. This may occur in deposits of monomeric IgG0 on cell surfaces such as the synovium or in aggregated IgGRFs. 4% of IgG in RA synovial fluid is in the form of soluble aggregates, and more are deposited on the synovium (Leader et al - personal communication). MBP is also present in RA synovial fluid (Malhotra et al 1995) and it is possible that the activation of complement through the MBP binding IgG0 may provide an additional route of complement activation within the rheumatoid joint.

5.4 Pure glycoforms of IgG0 may allow further functions of G0 sugars to be probed

Aglycosylated and degalactosylated IgG bind less efficiently to C1q and Fcγ receptor than normal IgG while protein A binding is not altered (Leatherbarrow et al 1985, Tsuchiya et al 1989). In murine IgG2b (and predicted in human IgG - Duncan et al 1988) the C1q binding motif involves Glu318, Lys320 and Lys322 (Duncan and Winter 1981), the Fcγ receptor binding site has been localised to Leu235 (Duncan et al 1988) and protein A requires an intact CH2/CH3 interface. These data suggest that, while the sugars do not affect the CH2/CH3

interface, they may maintain the integrity of the CH2 domains and the hinge in a way that allows C1q and Fcγ receptor to bind efficiently to IgG (For review: Brekke et al 1995). In another study (Leatherbarrow D.Phil) aglycosylated IgG was more susceptible to papain cleavage than normal IgG, but agalactosyl IgG has not been studied. In this chapter there was no correlation between the percentage of IgG0 (figure 21), G1 or G2 (data not shown) and the susceptibility of IgG to papain cleavage at Lys224/225. These data may suggest that this cleavage is not affected by the nature of the sugar at Asn 297. However, there are problems associated with interpreting these data. No information is available regarding the combinations of sugars which can be accommodated in the Fc, and there is no direct correlation between the percentage of G0 structures and the number of molecules of IgG containing IgG0. For IgG to be cleaved both heavy chains must be susceptible to the enzyme, and the rate of cleavage of IgG containing two different Fc sugars will depend on the rate of cleavage of the less susceptible chain. The preparation of IgG0 involved *Streptococcal 6646K* β-galactosidase (Kiyohara et al 1976) which, in contrast to other known galactosidases, cleaves galactose residues from the protected sugars in the IgG Fc. The technology is now available to prepare IgG0 to examine the papain cleavage and other properties of IgG. For example, in passive transfer experiments (Rademacher et al 1994), primed mice injected with 70% IgG0 from mice with collagen type II arthritis developed arthritis, thereby implicating IgG0 in the pathogenesis of the disease .

Pure glycoform populations have now been prepared for two glycoproteins, IgG and RNase B by enzymatic modifications of the natural populations, and a way may now be open to explore further the role of these and other glycoforms in modifying the properties of the proteins to which they are attached.

6. References to chapter 6:

Amzel, L.M. and Poljak, R.J. (1979) Ann. Rev. Biochem. 48 961-997 Three dimensional structure of immunoglobulins.

Bole, D.G., Hendershot, L.M. and Kearney, J.F. (1986) J. Cell Biol. 102 1558-1566 Post-translational Association of Immunoglobulin Heavy Chain Binding Protein with Nascent Heavy Chains in Nonsecreting and Secreting Hybridomas.

Brekke, O.H., Michaelson, T.E. and Sandlie, I. (1995) Immunology Today 16 85-90 The structural requirements for complement activation by IgG: does it hinge on the hinge?

Brown, P.B., Nardella, F.A., Mannik, M. (1982) Arthritis and Rheumatism 25 1101-1107 Human complement activation of self associated IgG RFs.

Carson, D.A., Chen, P.C. and Kipps, T. (1991) J. Clin. Invest. 87 379-383 New roles for rheumatoid factor.

Chang, S-C. (1992) D. Phil. Thesis, Oxford

Duncan, A.R. and Winter, G. (1988) Nature 332 738-740 The binding site for C1q on IgG.

Duncan, A.R., Woof, J., Partridge, L.J., Burton, D. and Winter, G. Nature (1988) 332 563-564 Localisation of the binding site for the human high-affinity Fc receptor on IgG.

Diesenhofer, J. (1981) Biochemistry 20 2361-2370 Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from *Staphylococcus Aurelius* at 2.9- and 2.8-A resolution.

Dwek, R.A., Sutton, B.J., Perkins, S.J. and Rademacher, T.W. (1984) *In* Molecular Variants of Protein Biosynthesis and Clinical Relevance (Eds. Campbell, P.N. and Phelps, C.) Biochem. Soc. Symp.123-136 Structure-function relationships in Immunoglobulins.

Elson, C.J., Thompson, S.J., Westacott, C.I. and Bhoola, K.D. (1992) Autoimmunity 13 327-331 Mediators of joint swelling and damage in rheumatoid arthritis and pristane induced arthritis.

Furukawa, K., Matsuta, K., Takeuchi, F., Kosuge, E., Miyamoto, T. and Kobata, A. (1989) International Immunology 2 105-112 Kinetic study of a galactosyltransferase in the B-cells of patients with RA.

Goochee, C.F., Gramer, M.J., Anderson, D.C., Bahr, J. and Rasmussen, J.R. (1992) Frontiers in Bioprocessing II 199-240 Eds. Todd, P., Sikdar, S.K. and Bier, M. Pub. American Chemical Society, Washington D.C. The oligosaccharides of glycoproteins: factors affecting their synthesis and their influence on glycoprotein properties.

Harada, T. (1987) Anal. Biochem. 164 374 Systematic fractionation of oligosaccharides of human immunoglobulins.

Haragai, M. Kitani A., Hara, M., Hirose, T., Norioka, K.I., Susuki, K. J. (1988) Rheumatol. 15 1616-1622 Rheumatoid adherent synovial cells produce B-cell differentiation factor activity neutralisable by antibody to B-cell stimulatory factor-2/Interleukin 6.

Hay, F.C, Nineham, L.J. and Roitt, I.M. (1975) BMJ 3 203.

Helenius, A. (1994) Mol. Biol.Cell 5 253-265 How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum.

Holmskov, U., Malhotra, R., Sim, R.B. and Jensenius, J.-C. (1994) Immunology Today 15, 67-74 Collectins: collagenous C-type lectins of the innate immune defence system.

Homans, S.W., Pastore, A., Dwek, R.A. and Rademacger, T.W. (1987) Biochemistry 26 6571-6578 Solution conformations of N-linked oligosaccharides.

Hase, S., Ibuki, T., and Ikenaka, T. (1984) J. Biochem. (Tokyo) 95 197-203 Re-examination of the pyrimidlamination used for fluoresent labelling of oligosaccharides and its application to glycoproteins.

Hoppe, H-J and Reid, K. (1994) Current Biology (Structure) 2 1129-1133 Trimeric C-type lectin domains in host defence.

Iobst, S.T., Wormald, M.R., Weiss. W.I., Dwek, R.A. and Drickamer, K. (1994) J. Biol. Chem. 269 15505-15511 Binding of sugar ligands to Ca^{2+} - dependent animal lectins.

Koide, N., Nose, M., and Muramatsu, T. (1977) BBRC 75 838 Recognition of IgG by Fc receptor and complement.

Jefferis, R., Weston, P.D. and Stanworth, D.R. (1968) Nature, 219, 646-649 Relationship between the papain sensitivity of human gamma G immunoglobulins and their heavy chain subclasses.

Jefferis, R., Lund, J., Mizutani, H., Nakagawa, H., Kawazoe, Y., Arata, Y. and Takahashi, N. (1990) Biochem. J. 268 529-537 A comparative study of the N-linked oligosaccharides of human IgG subclass proteins.

Kobata, A. (1990) Glycobiology 1 5-8 Function and pathology of the sugar chains on immunoglobulin G.

Kiyohara, T., Terao, T., Shioiri-Nakano, K. and Osawa, T. (1976) Biochem. J. 80 9-17 Purification and Characterisation of β–N-Acetyl hexosaminidases and β-galactosidase from Streptococcus 6646K.

Leatherbarrow, R.J., Rademacher, T.W., Dwek, R.A., Woof, M.J., Clark, A., Burton, D.R., Richardson, N., and Feinstein, A. (1985) Mol. Immunol. 22 407-415 Effector functions of a monoclonal aglycosylated mouse IgG2a: binding and activation of complement component C1 and interaction with human monocyte Fc receptor.

Leatherbarrow, R.J. and Dwek, R.A. (1984) Molecular Immunology 21 321-327 Binding of complement sub component C1q to mouse IgG1, IgG2a and IgG2b: A novel C1q binding assay.

Lee, S-O., Conolly, J.M., Ramirez-Soto, D. and Poretz, R. D. (1990) JBC 265 5833-5839 The polypeptide of Immunoglobulin G influences its galactosylation *in vivo.*

Lund, J., Takahashi, N., Pound, J., Goodall, M., Nakagawa, H. and Jefferis, R. (1995) FASEB J. In Press Oligosaccharide - protein interactions in IgG can modulate recognition by Fcγ receptors.

Matsushita, M. and Fujita, T. (1992) J. Expt. Med. 176 1497-1502 Activation of the classical complement pathway by mannose binding protein in association with a novel C1s-like serine protease.

Miyamoto, T. and Kobata, A. (1989) J. Rheum. 16 285-290 Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG

Mizouchi, T. (1982) J. Immunol. 129 2016-2021 Structural and numerical variations of the carbohydrate moiety of human immunoglobulin G.

Mizuochi, T., Hamako, J. and Titana, K. (1987) Archives of Biochemistry and Biophysics, 257(2), 387-394. Structures of the sugar chains of mouse Immunoglobulin G.

Nardella, F. A., Teller, D. and Mannik, M. (1981) J. Exp. Med., 154, 112-125 Studies on the antigenic determinants in the self association of IgG rheumatoid factor.

Nose, M. and Wigzell, H. (1983) Proc. Nat. Acad. Sci. USA 80 6632-6636 Biological significance of carbohydrate chains on monoclonal antibodies.

Padlan, E. (1991) Proc. Am. Soc. Microbiol., Washington D.C. Fc receptors and the Action of Antibodies.

Parekh, R.B. (1987) D.Phil. Thesis Oxford

Parekh, R.B., Dwek, R.A., Sutton, B.J., Fernandes, D.L., Leung, A., Stanworth, D., Rademacher, T. W., Mizuochi, T., Taniguchi, T., Matsuta, K., Takeuchi, F., Nagano, Y., Miyamoto, T. and Kobata, A. (1985) Nature 316, 452-457 Association of Rheumatoid Arthritis and Primary Osteoarthritis with changes in the glycosylation pattern of total serum IgG.

Parekh, R.B., Dwek, R.A., Isenberg, D., Roitt, I.M and Rademacher, T.W. (1988a) J. Exp. Med. 167 1731-1736 Age related galactosylation of N-linked oligosaccharides of human serum IgG.

Parekh, R.B., Isenberg, D., Ansell, B., Roitt, I.M, Dwek, R.A.,and Rademacher, T.W. (1988b) Lancet i, 966-969 Galactosylation of IgG associated oligosaccharides: reduction in patients with adult and juvenile onset rheumatoid arthritis and relation to disease.

Parekh, R.B., Isenberg, D., Rook, G., Roitt, I.M, Dwek, R.A. and Rademacher, T.W. (1989) J. of Autoimmunity 2 101-114 A comparative analysis of disease associated changes in the galactosylation of serum IgG.

Pope, R.M., Teller, D.C. and Mannik, M. (1975) J. Immunol. 1975 365-373 The molecular basis of self association of IgGRFs.

Rademacher, T.W., Homans, S.W., Parekh, R., Dwek, R.A. (1985) Genes and proteins in immunity - in honour of Prof. R.R. Porter - Biochem. Soc. Symp. No: 51, 131-148 (ed. J. Kay, M.A. Kerr, A.F. Williams and K.B.M. Reid) Immunoglobulin G as a glycoprotein.

Rademacher, T. W., Williams, P. and Dwek, R.A. (1994) Proc. Nat. Acad. Sci., USA, 91, 6123-6127 Agalactosyl glycoforms of IgG autoantibodies are pathogenic.

Rademacher, T.W., Parekh, R., Dwek, R.A., Isenberg, D., Rook, G., Axford, J.S. and Roitt, I. (1988) Springer Semin. in Immunopathol. 10 131-149. The role of IgG glycoforms in the Pathogenesis of Rheumatoid Arthritis.

Roitt, I.M., Dwek, R.A., Parekh, R., B., Rademacher, T.W., Alavi, A., Axford, J.S., Bodman, K.B., Bond, A., Cooke, A., Hay, F., Isenberg, D., Lydyard, P., Mackenzie, L., Rook, G.A.W., Smith, M. and Sumar, N. (1988) J. Autoimmunity 1 499-506 The role of antigen in autoimmune responses with special reference to changes in the carbohydrate structure of IgG during rheumatoid arthritis.

Rook, G.A.W., Steele, J., Brealey, R., Whyte, A., Isenberg, D., Sumar, N., Nelson, J.L., Bodman, K.B., Young, A., Roitt, I.M., Williams, P., Scragg, I., Edge, C.J., Arkwright, P.D., Ashford, D.A., Wormald, M., Rudd, P.M., Redman, C.W.G., Dwek, R.A. and Rademacher, T.W. (1991) J. Autoimmunity 4 779-794 Changes in IgG glycoforms are associated with Remission of Arthritis during Pregnancy.

Sin, T-W and Fletcher, D. (1978) Immunology 15 107-117 Interaction of human C1q with insoluble IgG aggregates

Sliwinski, A.J., and Zvaifler, N.J.(1970) J. Lab. Clin. Med. Aug. 304-310 *In vivo* synthesis of IgG by rheumatoid synovium.

Tomana, M., Schrohenloher, R.E., Bennet, P.H., del Puente, Koopman, W.J. Rheumatol. (1994) Int. 13 217-220 Occurrence of deficient galactosylation of serum IgG prior to the onset of rheumatoid arthritis.

Tsuchiya, N., Endo, T., Matsuta, K., Yoshinoya, S., Aikawa, T., Kosuge,E., Takeuchi, F., Miyamoto, T. and Kobata, A. (1989) J. Rheumatol. 16 285-290 Effects of galactose depletion from oligosaccharide chains on immunological activities of human IgG.

Tsuchiya, N., Endo,T., Matsuta, K., Yoshinoya, S., Takenchi, F., Nagano, Y., Shiota, M., Furukawa, K., Kochibe, N., Ho, K. and Kobata, A. (1994) J. Immunol. 151, 2237-1146 Detection of glycosylation abnormality in IgG using N-acetyl specific *Psathyrella velutina* lectin.

Weiss, W.I., Drickamer, K. and Hendrickson, W.A. (1992) Nature 360 127-134 Structure of a C-type mannose binding protein complexed with an oligosaccharide.

Weiss, W.I. and Drickamer, K. (1994) Current Biology (Structure) 2 1227-1240 Trimeric structure of a C-type mannose-binding protein.

Wirella, G., and Parkhouse, R.M.E. (1971) Immunochemistry, 8, 243-250 Papain sensitivity of heavy chain sub-classes in normal Human IgG and localisation of antigenic determinants for the subclasses.

van Zeben, D., Rook, G.A.W., Hazes, J.M.W., Zwinderman, A.H., Zhang, Y., Ghelani, S., Rademacher, T.W. and Breedveld, F.C. (1994) British Journal of Immunology 33 36-43 Early galactosylation of IgG is associated with a more progressive disease course in patients with rheumatoid arthritis: results of a follow-up study.

Chapter 7

Variable site occupancy controls the turn over rate of tissue plasminogen activator in the fibrin stimulated activation of plasminogen

Publications associated with this chapter:

1. Mori, K., Dwek, R.A., Downing, A.K., Opdenakker, G., and Rudd, P.M. (1995) J. Biol. Chem. 270 3261-3267 The activation of type 1 and type 2 plasminogen by type I and type II tissue plasminogen activator

2. Parekh, R.B., Dwek, R.A., Rudd, P.M., Thomas, J.R., Rademacher, T.W. Warren, T., Wun,T-C., Hebert, B., Reitz, B., Palmier, M., Ramabhadran, T. and Teimeir, D.C. (1989) Biochemistry 28 7670-7679 N-glycosylation and *in vitro* enzymatic activity of human recombinant tissue plasminogen activator expressed in chinese hamster ovary cells and a murine cell line.

3. Wittwer, A.J., Howard, S.C., Carr, L.S., Harakas, N.K., and Feder, J., Parekh, R.B.. Rudd, P.M., Dwek, R.A., and Rademacher, T.W. (1989) Biochemistry 28, 7662-7669 Effects of N-glycosylation on in vitro activity of Bowes Melanoma and human colon fibroblast derived tissue plasminogen activator.

4. Opdenakker, G., Rudd, P.M. Ponting, C.J. and Dwek, R.A. (1993) FASEB 14 1330-1337 Concepts and Principles of Glycobiology.

Acknowledgements:

1. Western blots and zymography of tPA were carried out by Professor Ghislain Opdenakker.
2. Activation of type I,II and D tPA by type 1 and 2 plasminogen was carried out in collaboration with Mr. Kazuya Mori who carried out the kinetic studies.
3. Molecular modelling was in collaboration with Dr. A. Kristina Downing and Dr. Christopher. Ponting.
4. The kinetics studies relating to C127 and CHO tPA pre-date the registration of this thesis and have been published (Parekh et al 1989); the data are included here for comparison and completeness.

Abbreviations: PA: plasminogen activator; uPA: urokinase plasminogen activator; tPA: tissue plasminogen activator; PAI: Plasminogen Activator inhibitor; A-chain: residues 1-275 of tPA; B-chain residues 276-527 of tPA; GnT: GlcNAc transferase; M: denotes sample was prepared in cell lines in Monsanto Co.; R: denotes sample was prepared in cell lines in the Rega Institute; BM: Bowes melanoma; HCF: Human Colon fibroblast; C127: Murine cell line C127; CHO: Chinese Hamster Ovary.

Abstract:

Glycoproteins generally consist of populations of glycosylated variants of a single polypeptide (glycoforms). The potential oligosaccharide pathways available to a protein are dictated by the cell in which it is expressed, and the final composition of the N-linked glycoform population results from the cell specific processing of a common precursor, $Glc_3Man_9GlcNAc_2$, by the biosynthetic glycosylation machinery within the individual cell. The consequences of cell specific glycosylation for the functional activity of an enzyme, tissue plasminogen activator (tPA) was explored by comparing the fibrin dependent and independent activity of tPA expressed in Bowes melanoma (BM) cells and a Human Colon Fibroblast (HCF) cell line. At a single time point the rate of fibrin stimulated production of plasmin from plasminogen by tPA from the BM cells was 1.36 times faster than the rate for the HCF derived tPA. The rate of production of plasmin from plasminogen in the absence of fibrin, which was linear with respect to time, was 1.7 times faster for the BM derived enzyme. Subsequently, variations in glycoform patterns and the consequences of variable site occupancy for lysine binding and functional activity were explored in tPA expressed in 8 different cell lines and fractionated at the protein level by lysine sepharose affinity chromatography. In BM cells type I tPA (3 occupied sites), type II tPA (2 occupied sites) and a minor fraction (D*), with high affinity for lysine and increased fibrin dependent and independent activity, were recovered. The minor fraction was found to contain a dimer of type II tPA which was designated type D tPA, and a molecular model was constructed suggesting a possible structure for the dimer. In an indirect amidolytic assay involving native human Glu-plasminogen and fibrin, type II tPA showed a 2-fold higher activity than type I. To explore the combinatorial effect of the variable glycosylation status of both tPA and plasminogen, kinetic constants for fibrin dependent plasminogen activation were determined for combinations of type I, II and D tPA with type 1 and 2 plasminogen. Within a four-fold range the fastest rate was achieved by the combination of type D* (type II + D) tPA and type 2 plasminogen. N-glycosylation of plasminogen increased the Km value for activation by all tPA variants, N-glycosylation of type I tPA at Asn-184 decreased the kcat (turnover) values for the fibrin dependent activation of plasminogen over type II tPA, while type D* tPA showed the highest turnover rate. In the presence of fibrinogen fragments, N-glycosylation of plasminogen at site 289 modulates the kinetics of association of enzyme and substrate, while N-glycosylation at site 184 on tPA modulates the turn-over rate of the enzyme.

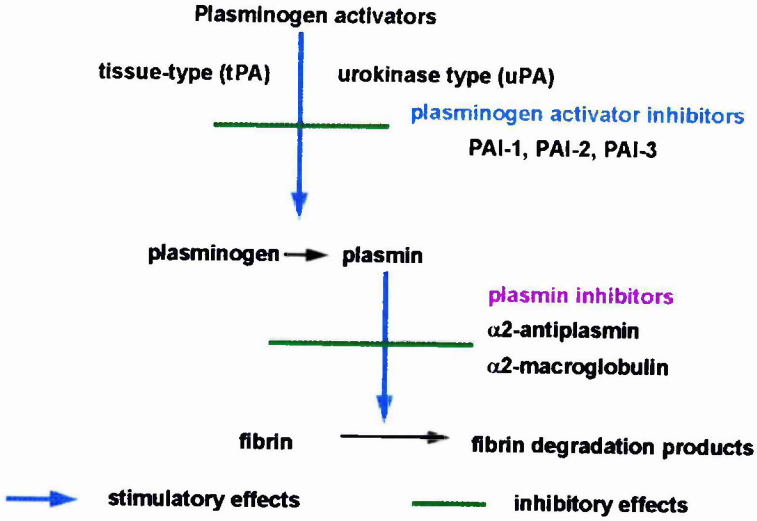


Figure 1: Fibrinolysis is regulated by a number of proteins which maintain a network of inhibitory and stimulatory reactions

Variable site occupancy controls the turn over rate of tissue plasminogen activator in the fibrin stimulated activation of plasminogen

1. Background

1.1 Plasminogen activators

Plasminogen activators (PA) are found in many tissues and fluids of the human body. They are proteolytic enzymes, whose only known substrate is the zymogen plasminogen. Plasminogen is cleaved by plasminogen activators to produce plasmin, a non-specific protease which is capable of degrading directly or indirectly most extracellular proteins (Vassalli et al 1991). Most of these enzymes are serine proteases which are modulated by several inhibitors, and are generated only in response to specific stimuli. Plasminogen activators include urokinase (uPA) and tissue-type plasminogen activator (t-PA), both of which are secreted by vascular endothelial cells. Plasmin has a broad spectrum of substrates and catalyses protein degradation in many biochemical systems including ovulation, neural development, embryogenesis, wound healing, tumour growth and metastasis, inflammation, extracellular matrix re-modelling and fibrinolysis (Dano et al 1985, Opdenakker et al 1992). Fibrinolysis and blood clotting are two dynamic processes which, under normal conditions, are in balance and disorders in either may lead to life threatening thrombotic or haemorrhagic events. In particular, fibrinolysis is regulated by a number of proteins which maintain a network of inhibitory and stimulatory reactions (Fig.1). The three main functional characteristics of tPA, which is present at 0.1 nM in serum, are binding to specific PA inhibitors (PAI), activation of plasminogen to plasmin, and specific binding to fibrin, leading to strongly enhanced plasminogen activation (Collen and Lijnen 1991).

1.2 Structure of tPA, plasminogen and fibrin

t-PA (70kD) was discovered by Astrup and Permin (1947) and was first purified from uterine tissue (Rijken et al 1979). However, tumor cell cultures have proved a more convenient source for studying the biochemical properties of tPA variants (Rijken et al 1981). It is present at only 0.1 nM in serum. The molecule has been well characterised at the protein level. It consists of 527 amino acids in five domains each of which show homology with domains of other proteins; these are the finger domain (homologous with a domain in fibronectin), growth factor domain (epidermal growth factor), kringle (K)1 and K2 domains (plasminogen and prothrombin), and a protease domain specific for serine proteases (Fig. 2). The inactivation of t-PA by its primary inhibitor, PAI type1, (PAI-1) is achieved through binding to the catalytic site, Asp371His322Ser478, (which is analogous to Asp194His40Ser32 in chymotrypsin and trypsinogen - Madison et al 1993) and to a secondary binding region, which includes the residues 296-304 while the finger domain and K2 are involved in fibrin binding (Horrevoets et al 1994). Two chain t-PA is derived from the single chain molecule by plasmin catalysed cleavage at Arg 275 (A-chain 1-275; B-chain 276-527); the molecule undergoes a major conformational change, but remains held together by disulphide bonds. The catalytic efficiency of the single

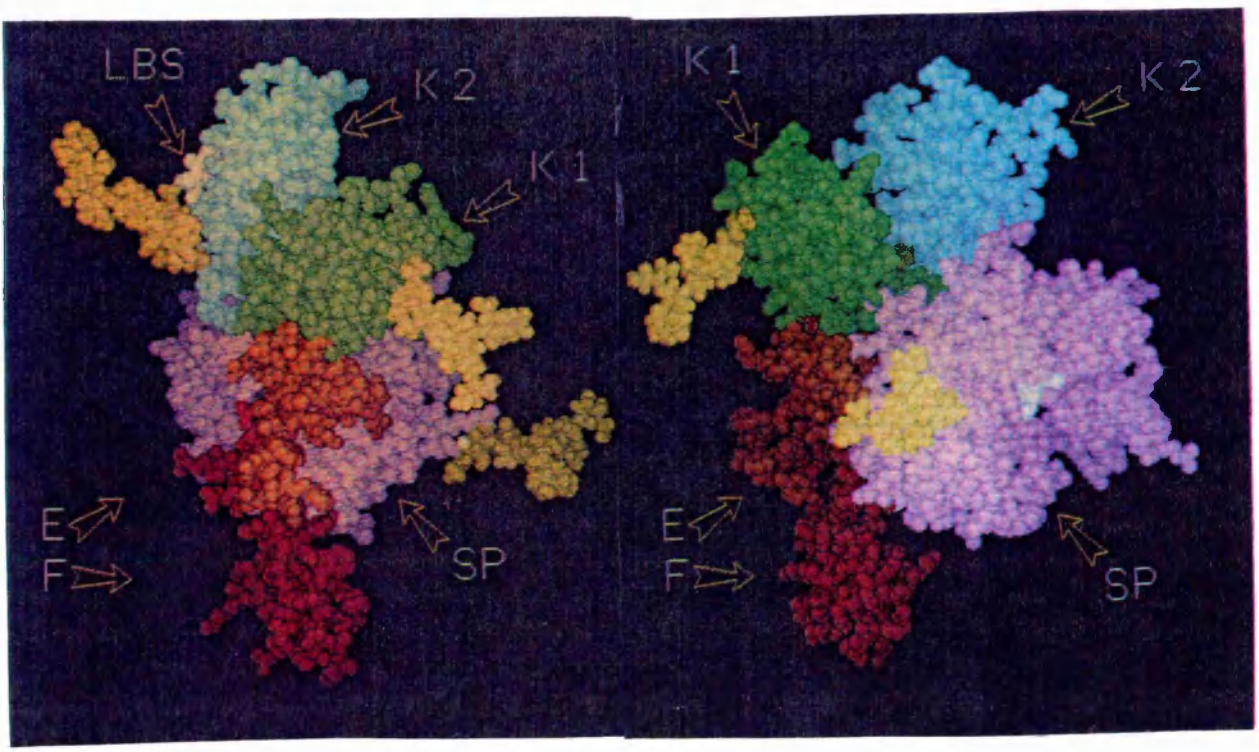


Figure 2: Two views of a model of tissue plasminogen activator *showing a hypothetical spatial arrangement of domains: the fibronectin type 1 'finger' (F) domain (shown in red), the epidermal growth factor-type (EGF) domain (brown), kringles 1 and 2 (green and blue, respectively) and the catalytic serine protease domain (purple). The F, EGF and serine protease domains have been modelled from high resolution structures (Baron et al 1990, Cooke et al 1987, Walter et al 1982); kringle 2 is the known crystallographic structure (de Vos et al 1992) from which the kringle 1 structure has been modelled. The figure illustrates the relative sizes of the sugars (shown in yellow) with respect to the protein for type I t-PA which is glycosylated at Asn 117, Asn184 and Asn 448. Type II t-PA lacks glycosylation at Asn 184. The complex bi-antennary oligosaccharide at Asn 184 appears to hinder the approach of macromolecules to the lysine-binding site (LBS) of kringle 2 (shown in white). The catalytic triad residues of the serine protease domain are also shown in white. (Opdenakker et al 1994)*

chain tPA zymogen is unusually high, approximately 15% of that of 2-chain tPA. (Boose et al 1989). t-PA contains four potential N-linked glycosylation sites, one of which (site 184) is variably occupied. In type I the sites in K1 (Asn117SerSer), K2 (Asn184GlySer), and in the protease domain (Asn448ArgThr) are all glycosylated, while in type II only sites 117 and 448 are occupied (Pohl et al 1984, Parekh et al 1989, Spellman et al 1989), the fourth sequon at Asn228ProSer, is not glycosylated. In addition, both type I and type II tPA contain an O-linked fucose residue at GlyGlyThr 61 (Harris et al 1991, Harris and Spellman 1993). In common with other glycoproteins, tPA shows species, cell and site specific patterns of glycosylation (Parekh et al 1989a; Wing, D.R. et al unpublished data). At Asn117 oligomannose sugars are highly conserved between species (Howard et al 1991), while the majority of the glycans at Asn184 and Asn448 are complex and hybrid structures. Occupancy of site 184 has been shown to affect the fine structure of the glycan population at site 448 demonstrating that glycosylation at one site can influence the processing at another. In Bowes melanoma type I tPA the major species at site 448 were neutral glycans of the complex or oligomannose type and only 13% of the glycans were sialylated or sulphated complex or hybrid structures. In type II, however 72% of the structures were sulphated complex glycans and there were relatively few neutral sugars (Parekh 1989a, Wing unpublished data). This suggests that in type I tPA the presence of a glycan at site 184 restricts the access of some of the glycosylation processing enzymes at site 448, reducing their efficiency.

tPA interacts with plasminogen, which is the pro-enzyme form of plasmin. Plasminogen is a single chain glycopeptide (M.Wt. 92kD) containing 5 kringle regions and a serine protease domain (Fig 3) and, like tPA, is a mixture of two major glycoforms (Brockway and Castellino 1972) which have the same amino acid sequence (Sottrup-Jensen et al 1978). Both glycoforms are O-glycosylated at Thr 345 and the sequence contains an N-glycosylation site at Asn289ArgThr which is occupied in type 1 (33%), but not in type 2 plasminogen (67%) (Hayes and Castellino 1979 a,b,c). The transition from plasminogen to active plasmin involves a cleavage of the Arg 561-Val 562 peptide bond in the serine protease domain after which a larger N-terminal A chain remains linked by two disulphide bonds to the smaller C-terminal B-chain which contains the protease active centre residues His 603, Asp 646, and Ser 741.

Fibrin, which forms a ternary complex with tPA and plasminogen, is derived from fibrinogen, which is a highly soluble plasma protein. Fibrinogen is converted into the insoluble monomer, fibrin, by the proteolytic action of thrombin which cleaves four Arg-Gly peptide bonds in fibrinogen, releasing fibrinopeptides and leaving fibrin monomers (Fig 4) which associate to form long insoluble fibres of fibrin.

1.3 Activation of plasminogen by t-PA within a ternary complex with fibrin.

t-PA activates plasminogen in both a fibrin dependent and independent manner. Under normal physiological conditions activation of plasminogen by tPA occurs slowly. However, in the presence of fibrin, to which both plasminogen and tPA bind, ternary complexes are formed and activation is enhanced by 2-3 orders of magnitude. The initial slow production of plasmin during

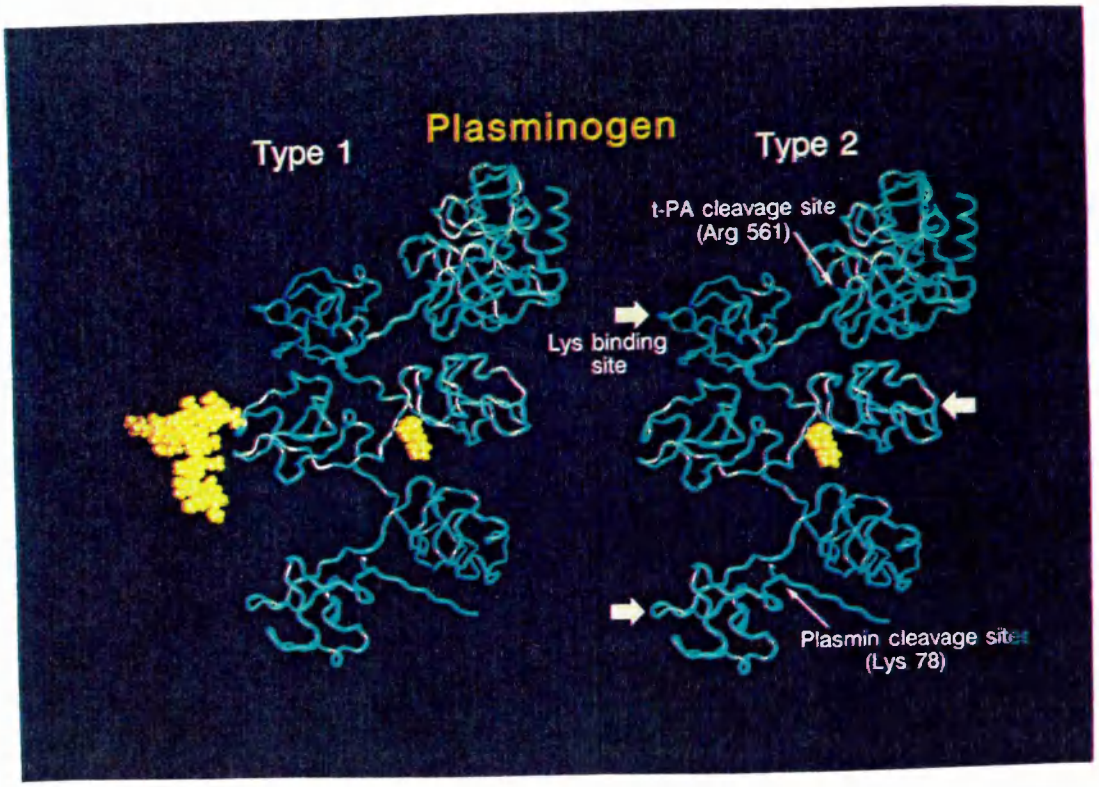


Figure 3: A schematic molecular model of plasminogen type 1 and type 2.

Plasminogen consists of five kringle regions and a serine protease domain. This model is constructed using the co-ordinates from the corresponding domains on tPA and is intended to convey the relative sizes of the oligosaccharides and the protein (Wormald, M. and Rudd, P. - unpublished data). Type 1 plasminogen (left) has two occupied glycosylation sites - at Asn289ArgThr in kringle 3 and at Thr345 in kringle 4. Type 2 lacks the N-linked sugar at Asn 289.

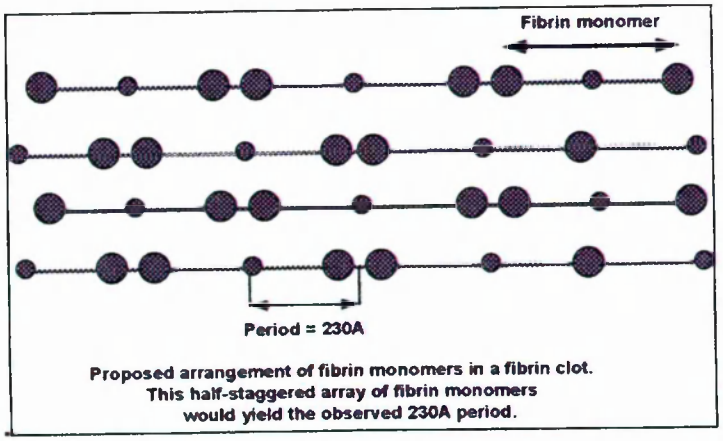


Figure 4: The arrangement of fibrin monomers in a fibrin clot

the "lag" phase observed in the fibrin dependent assay is almost entirely due to fibrin independent activation of plasminogen. As the reaction proceeds the rate of plasminogen activation increases for three reasons. First, as plasmin is produced the single chain t-PA is converted to the two chain form which degrades fibrin at 5 times the rate (van Zonneveld et al 1986). Second, during the degradation process carboxy terminal lysine residues are generated on fibrin, and this leads to increased binding of tPA and plasminogen. Third, during activation of plasminogen to plasmin, as a result of proteolytic degradation by plasmin, the full length native molecule, Glu-plasminogen (92kD) is converted to Lys-plasminogen (84kD). Amino acids are cleaved from the N terminus leaving a terminal lysine residue available binding tPA. In addition Lys plasminogen binds more tightly to fibrin than the full length molecule. The combination of all these processes leads to the formation of more stable ternary complexes and an exponential increase in rate.

Fibrin binding to t-PA involves the finger domain (von Zonneveld et al 1986) and K2 (Bennett et al 1991). High affinity binding of tPA to fibrin (Kd 3.2 +/- 0.6μM) is the result of co-operative binding between the two low affinity binding sites (Horrevoets et al 1994). The binding of plasminogen to fibrin is mediated by three lysine binding sites located in K1 and K4 and particularly K5 in the region of the molecule known as mini plasminogen (Val 422-Asn 790). The N-linked carbohydrate chain at Asn 289 decreases the binding of purified plasminogen fragments containing the high affinity lysine-binding site to fibrin by 85% (Lijnen al 1981); however, for intact Lys plasminogen, which binds more strongly to fibrin than Glu plasminogen, the difference between the two glycosylated variants is only minor. This suggests that other interactions apart from those involving the high affinity lysine binding site might contribute to the binding of plasminogen to fibrin.

1.4 The role of glycosylation in the functional activity of tissue plasminogen activator

The role of the carbohydrates in modulating the catalytic properties of tPA has been confirmed in many studies (Wittwer et al 1989, Parekh et al 1989, Berg et al 1993, Howard et al 1991). In these the fibrinolytic activity of type II tPA has consistently been found to exceed that of type I tPA (Einarrsson et al 1985), regardless of the cell line in which the tPA is produced. Variable occupancy of site 184 influences the rate at which plasmin is generated in two ways: the presence of an oligosaccharide at site 184 decreases the rate both by inhibiting the plasmin-mediated conversion of single to two chain tPA and by decreasing the affinity of tPA for lysine. In the first place, plasmin catalyses the cleavage of the Arg275-Ile bond in native single-chain tPA to form a disulphide bond linked two-chain species. The second order rate constant (kcat/Km) for type II tPA is about twice that for type I tPA (Wittwer and Howard 1990) suggesting that glycosylation at site 184 hinders the conversion of single- to two-chain tPA. Single chain tPA has a lower enzyme activity than the two-chain form (Ranby et al, 1982b Wallen et al 1983). Secondly, fibrin clot assays have shown that type II tPA binds more efficiently to fibrin (Wittwer et al 1989), moreover the decreased affinity of type I tPA for lysine is well established and affinity chromatography is commonly used to separate type I tPA from

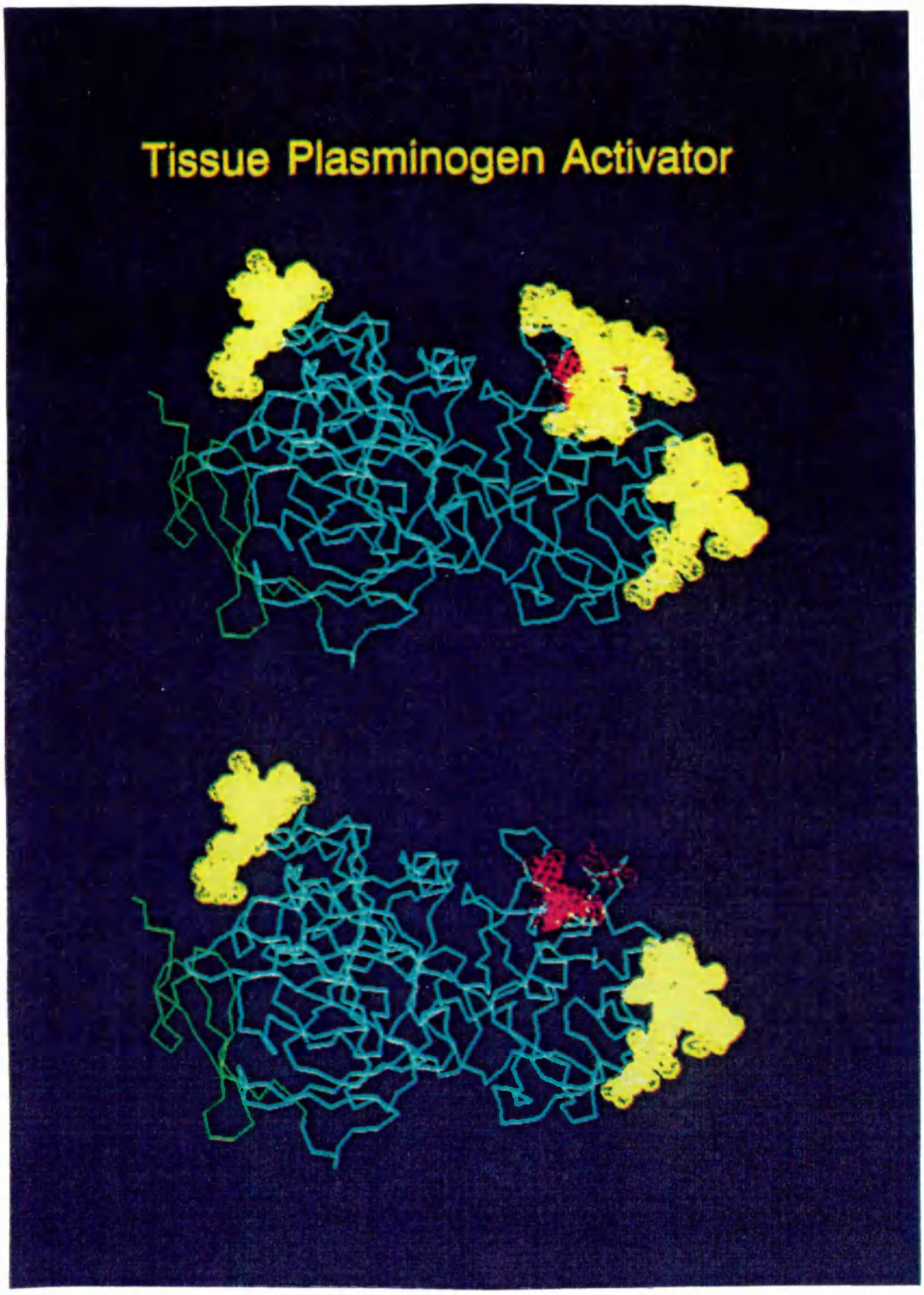


Figure 4b: The sugars at Asn 184 may sterically hinder a lysine binding site (red) on tPA type I *(above). This may inhibit the affinity of tPA for lysine and decrease the rate of fibrin stimulated activation of plasminogen compared with type II tPA (below).*

type II. These data suggest that the presence of a glycan at site 184 interferes with the binding of tPA to exposed lysine on degraded fibrin (Fig. 4b).

1.5 The role of tPA glycosylation in recognition

The carbohydrates also function as recognition markers and are involved in the clearance of tPA through receptors in the liver:

(i) Human tPA has been shown to be partly cleared through the macrophage mannose receptor located on liver endothelial cells and Kupffer cells (Owensby et al 1988, Krause et al 1990, Rijken et al 1990) and in rat liver cells by Smedsrod et al (1988). Deletion of site 117 by site directed mutagenesis (Hotchkiss 1988) or enzymatic removal of the oligomannose sugars using Endo-H (Tanswell et al 1989) both decreased the clearance rate.

(ii) The asialoglycoprotein receptor was shown to be involved in the hepatic clearance of tPA in rabbits (Cole et al 1993), and clearance of tPA by the mannose and galactose receptors in the liver was demonstrated in rats by Smedsrod et al (1988). Increased branching and desialylation of the oligosaccharides at sites 184 and 448 has been shown to increase the affinity of tPA for this receptor (Lee et al 1983).

(iii) Studies of a mutant tPA, in which site 117 was deleted (Cole et al 1993), suggested that glycosylation at site 184 exerts some steric or conformational effect which attenuates clearance of tPA through the peptide mediated clearance through the liver parenchymal cells in rabbits.

(iv) The conserved O-linked fucose residue at Thr 61 (Harris et al 1991, Harris and Spellman 1993) mediates tPA binding to Hep G2 cells *in vitro* through a calcium dependent mechanism after which tPA is rapidly internalised and degraded (Hajar and Reynolds 1993).

2. Introduction to chapter 7

The structural and functional aspects of the molecules described earlier in this thesis have illustrated many of the well established characteristics of protein glycosylation. In tPA many of these properties operate together in this complex, multidomain enzyme. For example, as in the CHO cell derived leucocyte antigens (ch.2), the processing of the N-linked glycans on tPA depends on the glycosylation machinery and the culture conditions in the cells in which the enzyme is expressed. Studies of the leucocyte antigens and of IgG (ch.6) indicated that the protein plays a role in determining its own glycosylation; in tPA, specific glycosylation at individual sites has been well established, and this illustrates both the importance of the local protein structure and the effect which glycosylation at one site can have on the processing at another. In common with RNase (ch.4), tPA contains a glycosylation site which is variably occupied, and in both enzymes the glycans shield substantial areas of the protein surface since they are large in comparison with the protein domains to which they are attached (in tPA the glycans are approximately the same size as a kringle domain). In RNase, the occupancy of Asn34 modulates the functional activity of the enzyme; in tPA, variable occupancy of site 184 modulates both the conversion of single to two chain tPA and the binding of tPA to fibrin and, as a consequence, the fibrin dependent activation of plasminogen to plasmin. The absence of a

sugar at site 184 allows tPA to bind to fibrin through K2 and the finger region in a conformation which allows optimal stimulation of plasminogen; the N-linked glycan situated close to the membrane in human erythrocyte CD59 may orient the binding site on the protein for optimal interaction with the C5-9 complex (ch.3). *In vitro* the agalactosyl sugars on IgGFc bind the mannose binding protein and the Fab glycans may be involved in immune complex formation, while in tPA, *in vivo*, glycans have been shown to be recognition markers for the clearance of tPA by the asialo glycoprotein receptor (Asn184 and Asn448), the macrophage mannose receptor (Asn117) and the hepatic receptor for fucose (Thr61).

In chapter 2 a study was made of five different proteins (CD2, CD5d1, CD5+CD4d3,4, CD48, and sCD59) all expressed in the same Chinese Hamster Ovary (CHO) cell line. In contrast, in this chapter, a study has been made of a single protein, tPA, which has been expressed in a number of different cell lines (Human colon fibroblast (HCF), Bowes melanoma (BM), CHO, murine C127 (C127) and MCF7). The role of glycosylation in modulating some of the functions and properties of t-PA has been explored from several angles. First, the effects of cell specific glycosylation on the overall rate of activation of plasminogen both in the absence and presence of fibrin has been explored in HCF and BM tPA. The influence of cell specific glycosylation and variable site occupancy on the affinity of each of the recombinant tPA glycoform subsets for lysine has been noted and exploited to prepare subsets of tPA glycoforms and their rates of plasminogen activation have been examined. As a result a subset of glycoforms with increased enzyme activity was identified and partially characterised. A study of the role of variable site occupancy in both tPA and plasminogen in modulating the formation and turnover rate of the ternary complex with fibrin raised the possibility that glycosylation may influence the rate of other enzymes in the fibrinolysis pathway. In preliminary work to explore this the glycosylation of neutrophil gelatinase type IV has been examined (ch.8). Throughout this study molecular modelling has been used to gain insight into the steric implications of the glycosylation of tPA for the structure and functional activity of the enzyme.

3. Materials and Methods

3.1 Expression and purification of tissue plasminogen activator.

The therapeutic potential, as well as its value as a model system for studying glycosylation, has led to the establishment of a number of cell lines for the production of t-PA. 2-chain t-PA prepared from the following cell lines was supplied through collaborations with Dr. J. Feder (Monsanto Company (M)) and Prof. G. Opdenakker (Dept. of Microbiology and Immunology, Rega Institute, University of Leuven, Belgium (R)): Bowes Melanoma (BM), Human Colon Fibroblast (HCF) from serum containing and serum free media, murine (C127), Chinese Hamster Ovary (CHO), Phorbol Ester treated Melanoma cells, β-estradiol treated MCF-7 cells, and unstimulated Bowes Melanoma (BM).

3.2 Preparation of the glycosylated variants of t-PA from different cell lines by lysine - sepharose chromatography

800 ug of t-PA at 1 mg/ml in 1M ammonium bicarbonate were diluted with 1.2 ml of buffer, (10mM sodium phosphate buffer pH 8.0 containing 0.15M KSCN 0.001% Tween 80, and 25 KIU of aprotinin/ml), and applied to a lysine sepharose column (45 x 0.9 cm) equilibrated in the same buffer. The column was washed with 35 ml of equilibration buffer, at a flow rate of 5.5ml/h and eluted with a linear gradient of 0 - 0.25M L-arginine in the same buffer (160ml); 1.0ml fractions were collected. The column eluate was monitored by the direct amidolytic assay using the artificial substrate S2288 (Kabi Vitrum). Aliquots were assayed in 96-well microtitre plates using either a Molecular Devices kinetic microplate reader or a Titertek single time point plate reader. Fractions were pooled as required.

3.3 Preparation of fibrinogen fragments: Soluble fibrin fragments were produced by cleavage of human fibrinogen with cyanogen bromide, as described by Verheijen et al (1983). 100mg fibrinogen (M.Wt. 340kd) + 0.13mg cyanogen bromide in 10 ml formic acid were incubated overnight at room temperature. The resulting fragments were dialysed against water and stored at -20°C. The concentration equivalents of fibrin fragments were estimated by assuming that one fibrin fragment is generated from each fibrinogen molecule i.e. that 1 mole fibrinogen is equivalent to 1 mole of fibrin fragments. On this basis the final solution was calculated to be 29.4μmolar.

3.4 Assays of t-PA activity

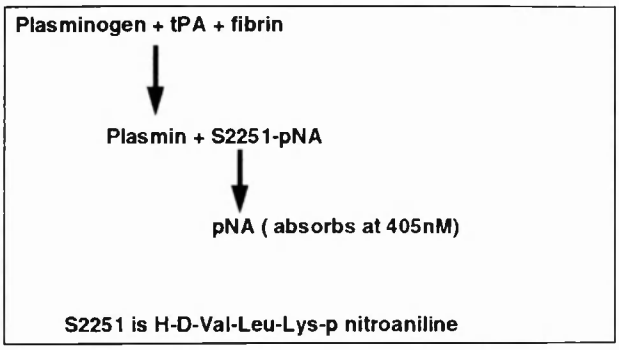
a. Direct Amidolytic Assay measuring protease activity.

The direct amidolytic assay was performed at 37°C using the synthetic substrate S-2288 (Kabi Vitrum) in a reaction mixture consisting of 1 mM S-2288 in 0.1M Tris buffer, pH 7.5 (at 25°C) containing 0.1% Tween 20 (v/v), and the absorbance monitored at 405nm. S-2288: (H-D-Isoleucyl-L-prolyl-L-arginine-p-nitroanilide-dihydrochloride; KabiVitrum) is a chromogenic substrate sensitive to a broad spectrum of serine proteases. S2288 binds to tPA at the catalytic triad in the active site in the protease domain where it is cleaved, releasing *p*-nitroaniline which absorbs at 405nm.

b. Stimulated Indirect Amidolytic Assay.

This was performed at 37°C in a reaction mixture consisting of 0.13μM Glu -type human plasminogen (Kabi Vitrum), 0.7mM plasmin substrate S-2251 (Kabi Vitrum), 0.6 μM soluble fragments of human fibrinogen or fibrin, and t-PA of defined protein concentration (determined by amino acid sequencing) or protease activity (determined by the direct amidolytic assay) in 0.1M Tris buffer, pH 7.5, containing 0.1% Tween 20 and monitored at 405 nm on a Pye Unicam 8620 spectrophotometer. The rate of conversion of plasminogen to plasmin was determined from plots of absorbance (A) at 405nm versus reaction time and also from derivative plots of

dA /dt versus t^2. The unstimulated indirect assay was performed in the same way without the addition of fibrin. The reaction is shown schematically below:



3.5 SDS PAGE and Immunoblotting. SDS polyacrylamide gel electrophoresis was performed according to the standard method. After electrophoresis, the separated samples in the gel were transferred to blotting membrane (Immobilon or nitrocellulose, Millipore) by electroblotting. The tPA bands were developed using rabbit anti-tPA polyclonal antibody and alkaline phosphatase or peroxidase conjugated to anti-rabbit IgG (Opdenakker et al 1988). Casein was used as the plasmin substrate in the direct zymographic analysis (Roche et al 1993).

4. Results

4.1 Determination of optimal conditions for the indirect stimulated assay

4.1.1 t-PA is not stimulated by fibrinogen

BM and HCF t-PA (0.142nmolar) were incubated with increasing amounts of intact fibrinogen (0 - 0.586 M) under the conditions for the stimulated indirect amidolytic assay. Following the method of Verheijen et al (1982) which uses the kinetics derived by Drapier et al (1979) plots of ΔOD vs t^2 were made for each concentration of fibrinogen (data not shown). From these data the dependence of t-PA activity on fibrinogen (Fig 5a,b) concentration was plotted at a single time point (10min). The result demonstrates that t-PA is not stimulated by intact fibrinogen. The standard mixture of soluble fibrin fragments used in the following experiments was prepared by CNBr cleavage of fibrinogen, and this result establishes that residual uncleaved fibrinogen in the fibrin fragment preparation does not take part in the reactions.

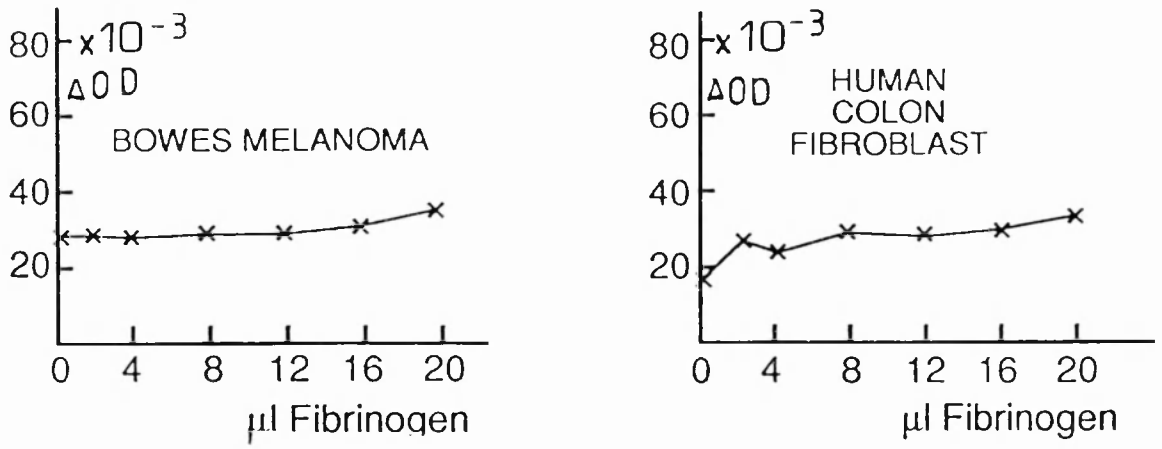


Figure 5a,b: Fibrinogen does not stimulate activity in either (a) Bowes melanoma (BM) or (b) Human Colon Fibroblast (HCF) tPA

4.1.2 Stimulation of t-PA depends on the concentration of fibrin fragments

Soluble fibrin fragments were prepared by CNBr cleavage of human fibrinogen. BM and HCF t-PA were incubated with increasing amounts of the fragments (0 - 0.586 M) under the conditions for the stimulated indirect amidolytic assay, keeping the concentration of plasminogen constant (0.13μM). Plots of ΔOD vs t^2 were made for each concentration of fibrin (Fig. 6a,b). From these data the dependence of t-PA activity on fibrin concentration at a single time point (10min) was plotted (Fig 6c). At concentrations higher than 0.353μmolar (12μl of standard solution) problems developed with turbidity and solubility in some samples, therefore saturating levels of fibrin could not be used. The optimal concentration of fibrin was 0.353μmolar, and this was used in subsequent assays. In this plot the y-axis represents the change in OD over the first 10 minutes of the reaction or the initial rate. Therefore Michaelis-Menton type analyses can be used for the interaction of constant concentrations of tPA and plasminogen in the presence of increasing concentrations of fibrin. At each fibrin concentration Bowes tPA was more active than HCF. The limiting factor for ΔOD after 10 minutes, "Vmax", was estimated to be 0.07 and 0.05 ΔOD/min respectively. The "Km" values for BM tPA were similar: 0.187μmolar and 0.192μmolar for HCF tPA. The amino acid sequences of BM and HCF tPA were identical and the differences between the two enzymes reside in the carbohydrate. Therefore these data suggest that differences in the glycosylation of tPA do not affect the formation of the ternary complex involving tPA, fibrin and plasminogen, but do influence the turnover rate of tPA.

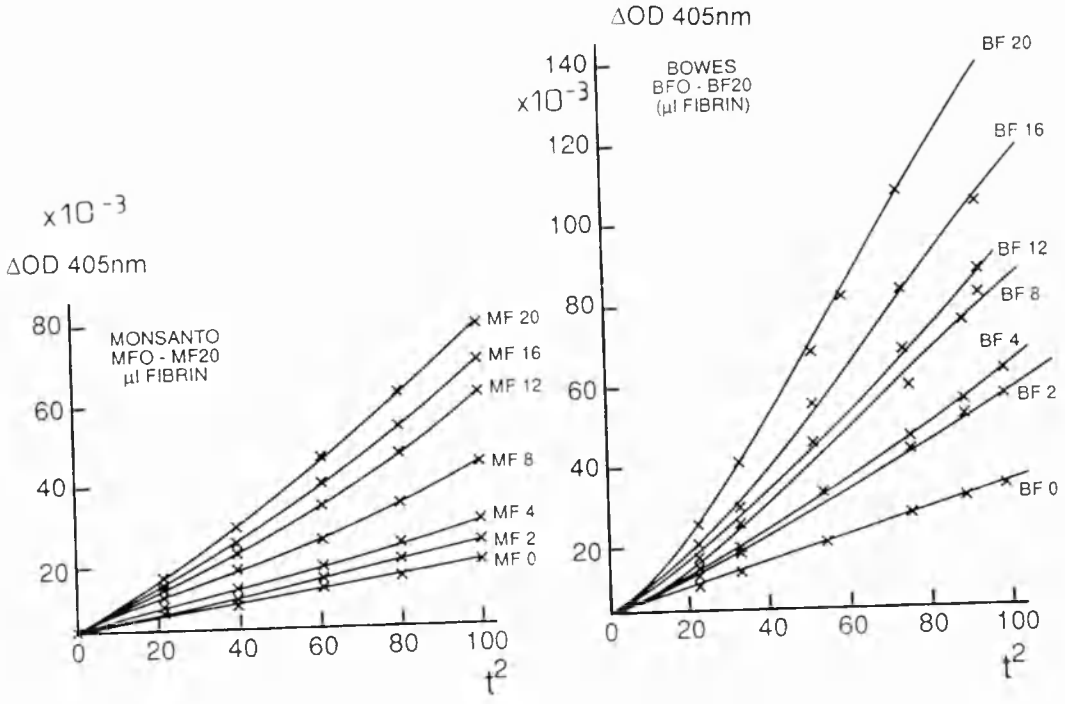


Figure 6a-b: ΔOD vs t^2 plots of activity of tPA (0.142nM) towards plasminogen (0.13μM), with increasing concentrations of fibrin (0-0.586M).

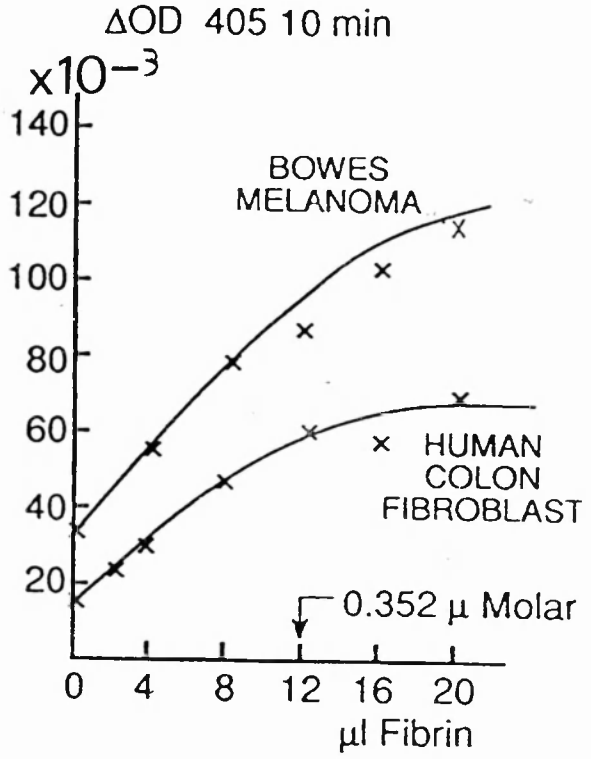


Figure 6c: t-PA activity depends on the concentration of fibrin fragments.

4.1.3 t-PA activity depends on the concentration of plasminogen

BM and HCF t-PA were incubated with increasing amounts of plasminogen ranging from 0 - 0.585μmolar using the previously determined concentration of fibrin (0.353μmolar) and 1.3nmolar t-PA. The rate of plasmin formation was measured by the indirect amidolytic assay using S2251 (Fig 7). "Vmax" and "Km" were determined for both types of tPA. When 50% saturating fibrin levels were used, 50% saturating plasminogen levels were reached at approximately 0.13μmolar (0.352 units), and this concentration of plasminogen was used in the subsequent assays. Although the "Km" values were similar (0.13μmolar for BM and 0.135μmolar for HCF), the "Vmax" for BM tPA (0.2 ΔOD/min) was more than twice as fast as that for HCF (0.095 ΔOD/min). This is consistent with the previous data (Fig.6c) which suggests that differences in the glycosylation of tPA influence the turnover rate but not the formation of the ternary complex (Fig. 7).

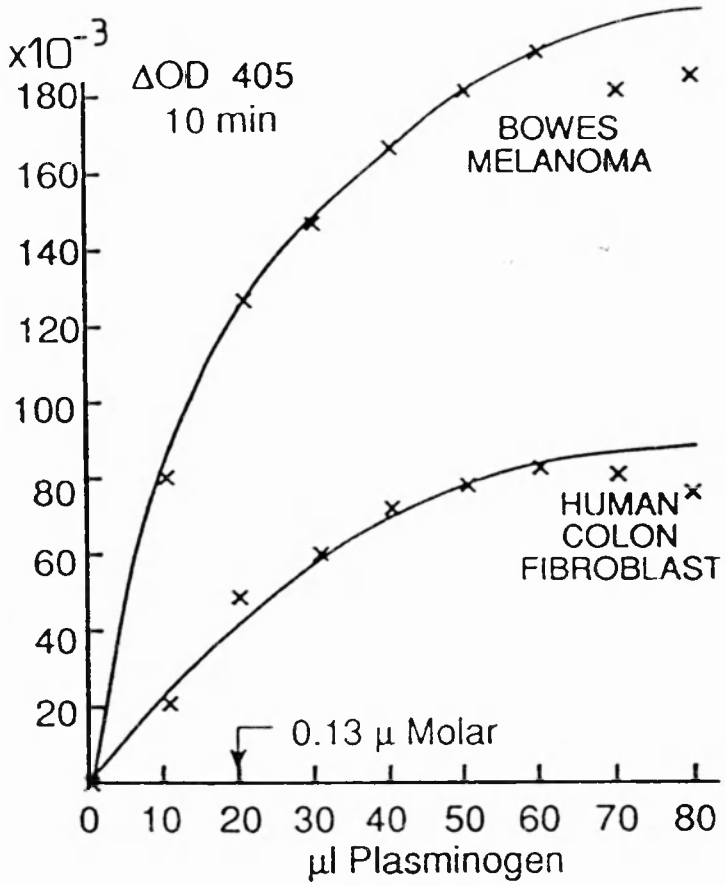


Figure 7: t-PA activity depends on plasminogen concentration.

In the next section the fibrin independent activities of BM and HCF tPA were compared using the optimised assay conditions.

4.2 Comparison of the activities of Bowes Melanoma and Human Colon Fibroblast tPA

4.2.1 The fibrin independent activity of BM tPA is greater than that of HCF t-PA.

The indirect assay was performed in the absence of fibrin to compare the unstimulated activity of BM and HCF t-PA towards plasminogen. Excess t-PA (14.2nMolar) was used since, as expected, the standard amount which had been optimised in the fibrin stimulated assay gave low activity in this assay over the time scale. There was no lag phase in the fibrin independent assay, and the rate of plasmin production was linear with respect to time^2 (Fig 8). The slope for BM t-PA was 0.034 ΔOD/min^2 compared with 0.02 ΔOD/min^2 for HCF. Fibrin independent activity is an important parameter because high levels of fibrin independent activity can lead to internal bleeding when recombinant tPA is used pharmacologically to disperse blood clots.

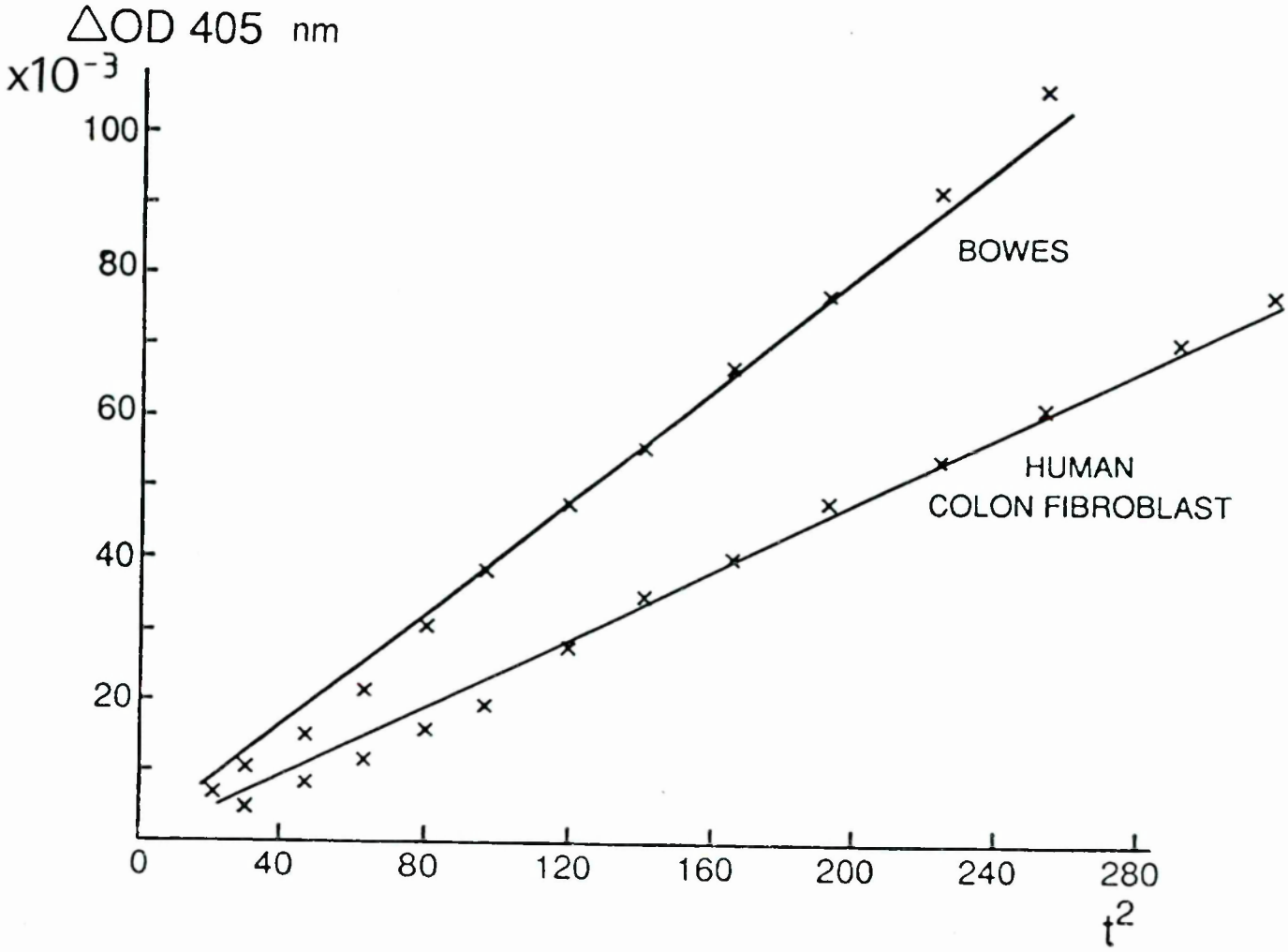


Figure 8: The rate of fibrin independent activity of BM and HCF tPA is linear with respect to time^2.

4.2.2 The combined fibrin independent and dependent activity of Bowes melanoma and Human Colon Fibroblast t-PA.

The total activity of tPA in the fibrin dependent activity includes a contribution that is derived from the fibrin independent activity. In this assay t-PA was added in increasing amounts (0-350 ng) to the reaction mixture for the indirect amidolytic assay. The concentrations of the fibrin (0.352μmolar) and plasminogen (0.13μmolar) were the optimal concentrations established above. The constituents of the reaction mixture were all pre-incubated to 37°C before mixing. The rate of change of absorbance at 405nm was measured using a parallel kinetics spectrophotometer. Plots of ΔOD vs time^2 (t^2) were drawn (data not shown). At t = 7min (t^2 = 49min^2) ΔOD was read from these plots and graphs of ΔOD at 7 min vs t-PA concentration were drawn (Fig 9). After 7 minutes the ratio of the rate of change of absorbance (i.e plasmin production) of BM vs HCF tPA when the tPA concentration was 2.48 nmolar was 1.36:1.

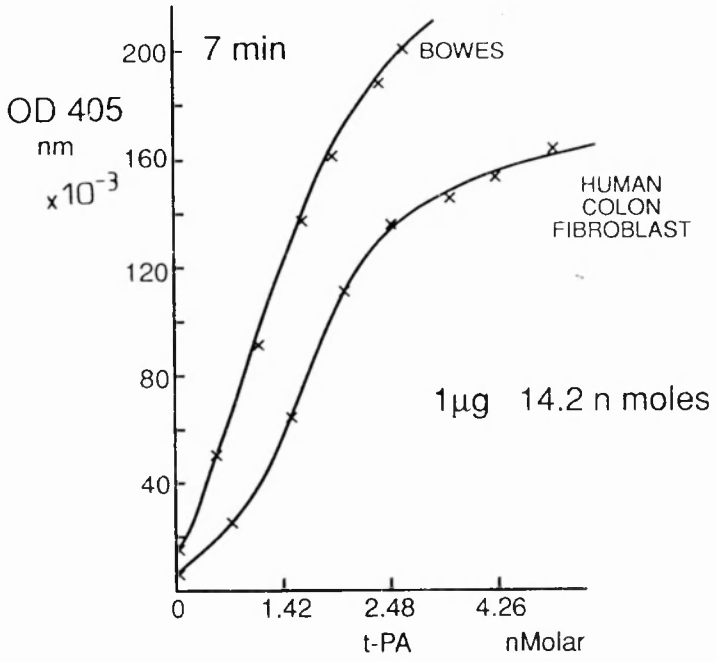


Figure 9: A comparison of the combined fibrin stimulated and unstimulated activities of BM and HCF tPA *(Ratios BM:HCF 1.36:1)*

4.2.3 Comparison of the fibrin dependent activity of Bowes and Human Colon Fibroblast t-PA.

The fibrin dependent activity of tPA alone was obtained by reading the absorbances in the indirect stimulated assay against a blank which contained everything except fibrin. In this way the fibrin independent activity generated in the sample was accounted for in the blank.

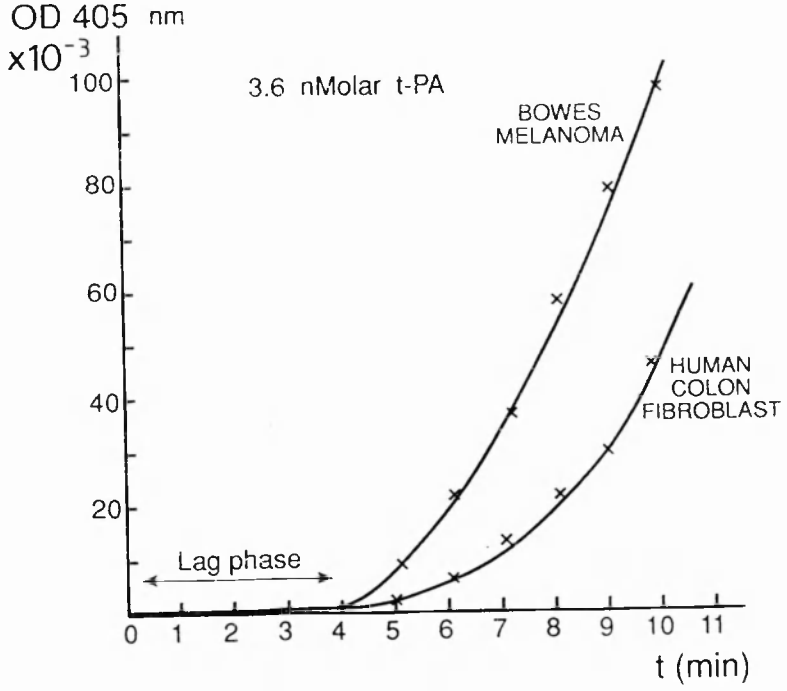


Figure 10: Fibrin dependent activity of BM and HCF tPA. *The contribution of the fibrin independent activity has been subtracted.*

In the indirect assay measuring both fibrin dependent and independent activity simultaneously, 7min from the start of the reaction, at a concentration of 2.48nMolar, BM t-PA liberated approximately 33% more plasmin from plasminogen than HCF t-PA at the same concentration (Fig. 9). In an assay measuring fibrin dependent activity alone the lag phase for both samples was about 4 minutes (Fig.10); however the initial rate of fibrin dependent cleavage of plasminogen by BM t-PA was 2.4 times as fast as the HCF t-PA. In this experiment no activity was observed in the lag phase since this was counteracted by reading the absorbances against a fibrin negative blank. These data indicated that overall the activity of BM tPA, which contains a higher proportion of less processed oligomannose structures, was greater than that of HCF tPA, and suggest that the differences in glycosylation which result from the expression of tPA in different cell lines are involved in modulating the activity of tPA. In the next section the effects of cell specific glycosylation on the glycoform pattern, the binding to lysine and the fibrin stimulated activity of subsets of tPA from a number of cell lines were investigated.

4.3 Cell specific glycosylation of t-PA.

t-PA was expressed in the following cell lines in two laboratories: (a) (from Professor G. Opdenakker): unstimulated BM, HCF from serum containing and serum free media, Phorbol Ester treated BM, β-estradiol treated MCF-7 (b) (from Monsanto Ltd.): C127, CHO, HCF and BM.

4.3.1 Profiling the glycoform populations of recombinant tPA from a range of cell lines:

t-PA was fractionated on the basis of the differences in the affinity of the glycosylated variants for lysine sepharose. Each sample of tPA (50μg) was applied to a lysine sepharose affinity column and eluted with an arginine gradient (0-0.5M). The fractions were assayed for protease activity in the direct assay using S2288, and characteristic profiles of the glycoform populations were obtained for each cell line (Fig. 11a-i). These reflect differences in the glycosylation of t-PA which result both from the different arrays of processing enzymes in each of the cell types and from variations in cell culture conditions. This profiling technique allows the composition of different populations of t-PA glycoforms to be compared at the protein level, and for rapid screening purposes this eliminates the need to release and analyse the carbohydrates. It also represents the distribution of glycoforms; this information is lost in oligosaccharide profiling using released sugars. The fractions eluted from the column were pooled as shown in figure 11. The pools contain sets of tPA glycoforms which have been separated on the basis of their different affinities for lysine. Those with lower affinity for lysine, and by implication for terminal lysine residues on degraded fibrin (Type I tPA), elute early in the concentration gradient, while the glycoforms with higher affinity (Type II tPA) elute later. Up to 5 sets of fractions from each tPA sample were collected (Fig.11,12 table 1). In each case the first pool contained a subset of glycoforms (type Ia) eluting in the conductivity range of 20.95-22.35ms/cm (within a band covering 1.4ms/cm), the second pool (type I tPA) and eluted between 21.63 and 23.68ms/cm (a band of 2.05) ms/cm, the third (which was apparent only in BM (R)) contained type III and type

Figure 11a-i: tPA from a number of cell lines was fractionated by lysine-sepharose chromatography. *Activity in each fraction (monitored by the direct amidolytic assay) was plotted against the fraction number. The conductivity of each fraction was also measured and the profiles are in line with respect to conductivity. The conductivity trace is shown in 11a and b and the conductivity values of the peak fractions are recorded in Table1.*

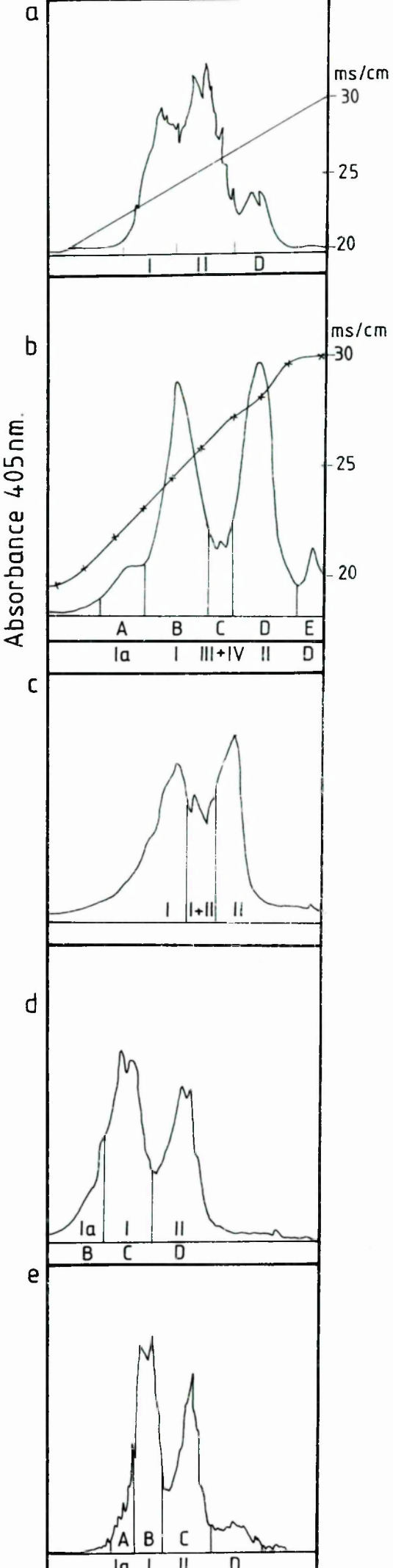
Human Colon Fibroblast (M)

Bowes Melanoma (M)

Bowes Melanoma (R)

Human Colon Fibroblast (R)

Murine C127 (M) Lot 1.



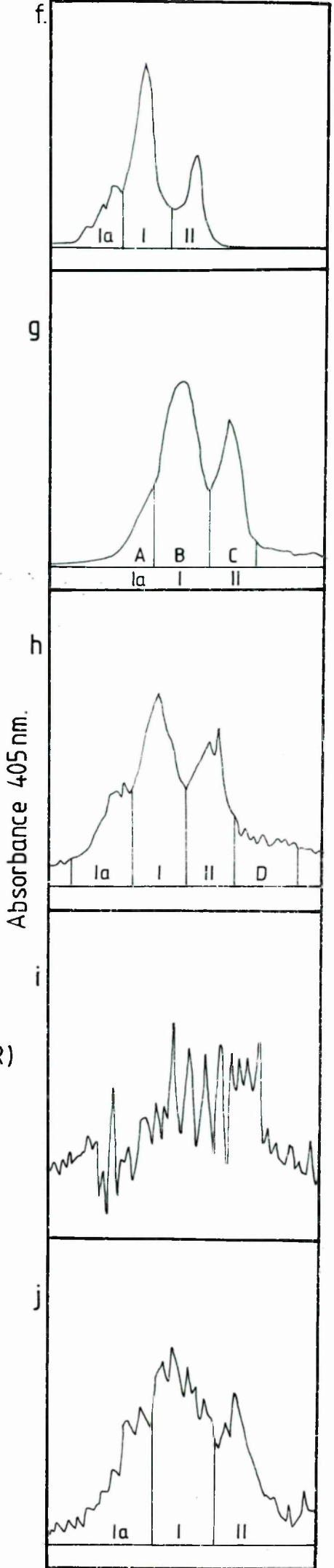
Murine C127(M) Lot 2.

Chinese Hamster Ovary (M)

Phorbol ester treated
Bowes Melanoma (R)

β-estradiol treated MCF7 (R)

Unstimulated
Bowes Melanoma (R)



IV tPA and eluted at 23.71ms/cm, Type II eluted in a conductivity band between 23.29 and 26.01ms/cm (band of 2.72) ms/cm and type D eluted between 25.28 and 26.04ms/cm (band of 0.76ms/cm). It is apparent that these conductivity bands overlap to some extent, suggesting that the lower affinity of type I and type II for lysine is not compensated for by alterations in the glycosylation. Some cell lines express type Ia tPA which elutes within the conductivity range of type I and some type II glycoforms elute in the type D range. However, interestingly, there is no overlap between type I and type II. The width of each conductivity band may be related to the heterogeneity of the glycoforms in the range of samples - type II is the broadest and most heterogeneous followed by type I>type Ia>type D. This suggest that type Ia and type D may be specific subsets of type I and type II tPA respectively.

In particular it was noted in this analysis that all the types of BM tPA, which was shown in the previous section to be more active in both fibrin dependent and independent assays than HCF tPA, eluted later in the gradient than the corresponding HCF glycoforms and contained a higher proportion of type Ia glycoforms.

tPA	type Ia	type I	type I+II	type II	type D
(a) HCFM	22.0	22.32		23.29	25.6
(b) BMM	22.12	23.68		26.01	
(c) BM(u:R)	22.08	22.94	23.71	25.29	26.04
(d) HCFR	20.95	21.63		23.87	
(e) C127	21.19	22.79		24.78	25.28
(f) CHO	22.47	23.57		25.17	
(g) PMA BM	21.43	22.93		25.03	25.66
(h) MCF7	21.62	23.57		25.55	
(i) uBM(R)	22.35	23.52		25.4	

Table 1: Conductivity at peak fraction for pools of glycoforms from a range of tPA samples eluted from a lysine sepharose affinity column. The units are ms/cm; error +/- 1%.

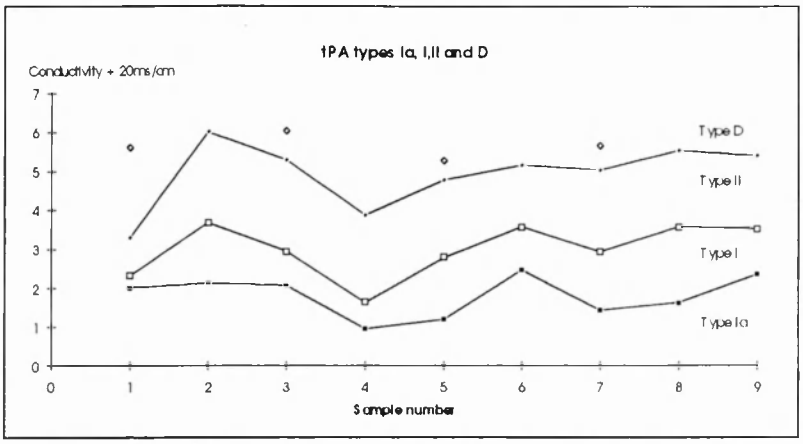


Figure 12: Conductivity of types Ia, I, II and D from a range of cell lines. *The data was taken from table 1.*

1: HCF (M) 2:BM (M) 3. uBM (R) batch 1 4. HCF (R) (serum in medium)
5. C127 6. CHO 7. PMA BM (R) 8. MCF7 (R) 9. uBM (R) batch 2
u = unstimulated

The glycoform subsets from figure 11a-i were analysed further by a combination of SDS PAGE, Western blotting and assays of fibrin stimulated activity. Some representative results are shown in the following section.

Unstimulated Bowes melanoma t-PA (Fig. 11c):

Immunoblot analyses of the glycoforms of BM t-PA from an unstimulated cell line, fractionated on lysine-sepharose, are shown in figure 13a. In contrast to the phorbol ester stimulated BM tPA, pool C contained two novel forms of t-PA migrating as a doublet with an apparent molecular weight of about 60kD. These have been noted in earlier studies of BM tPA (Opdenakker et al 1988) and have been designated Type III and Type IV tPA. They have not been investigated further, but the data indicate that t-PA from mammalian cell lines is more heterogeneous than was previously believed. Possible explanations for the existence of these structural variants include glycosylation changes, however the possibility of N-terminal, C-terminal or internal clipping or other post-translational modifications cannot be discounted. In addition, lane E, which contains material from the pool with the greatest affinity for lysine, pool IIb, contains Type II t-PA as expected, and also a component at 115 -125kD which reacts with the α-tPA antibody. This was designated type D tPA. Zymography of the same set of glycoforms (Fig.13b), using casein as the indicator substrate, resolved the BM t-PA sub-types from pools Ia, I, I+II, II and IIb, as well as type D tPA in pool IIb. The gel indicates that type D tPA, migrating at approximately 120kD, also has activity towards casein. The data also demonstrate that the proteins, which are re-natured in the gel before caseinolysis, have retained proteolytic activity.

The figure legends opposite should read:

Figure 13a: Zymography of tPA from pools Ia, I, I+II, II and IIb (Fig.11c), using casein as the indicator substrate. *Lanes A,B,C,D,and E are from blots of material from pools Ia, I, I+II, II, IIb (as in the immunoblot in figure 13a) respectively. The gel indicates that type D tPA, migrating at approximately 120kD, also has activity towards casein.*

Figure 13b: Analysis of glycoform populations of BM tPA by SDS PAGE, and immunoblotting. Lanes A,B,C,D,and E are blots of material from pools Ia, I, I+II, II, IIb respectively. Track F contains a sample of standard Bowes melanoma tPA for comparison.

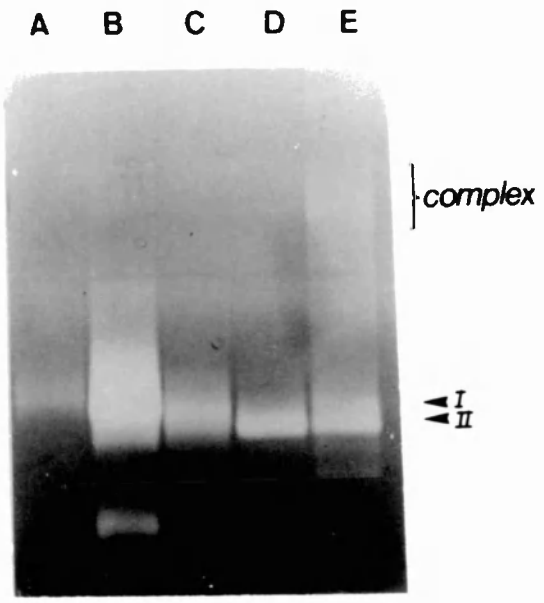


Figure 13a: Analysis of glycoform populations of BM tPA by SDS PAGE, and immunoblotting. Lanes A,B,C,D,and E are blots of material from pools Ia, I, I+II, II, IIb respectively.

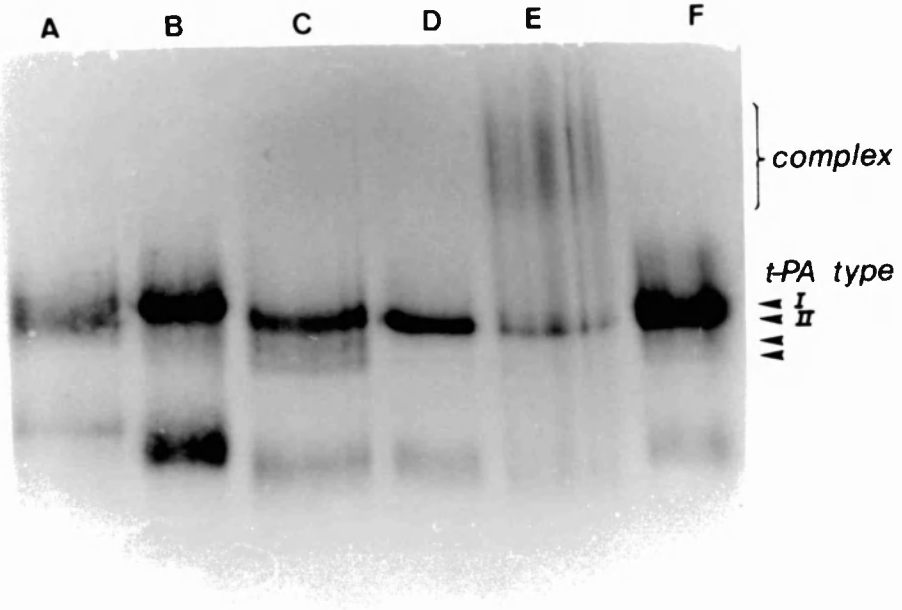


Figure 13b: Zymography of tPA from pools Ia, I, I+II, II and IIb (Fig.11c), using casein as the indicator substrate. *Lanes A,B,C,D,and E are from blots of material from pools Ia, I, I+II, II, IIb (as in the immunoblot in figure 13a) respectively. The gel indicates that type D tPA, migrating at approximately 120kD, also has activity towards casein.*

Human colon fibroblast t-PA (Fig.11d):

Immunoblot analyses of t-PA from the pools designated unbound, I, II and III (lanes A-D respectively) which were eluted from lysine-sepharose are shown in figure 14. Standard melanoma preparations, which contain t-PA Types I-IV, (lanes E and F) are included as controls. Mainly Type I and Type II tPA were detected in this gel. This may suggest that the levels of other tPA variants in this sample are too low to be detected or that the resolution of the gel does not allow a further separation of tPA variants.

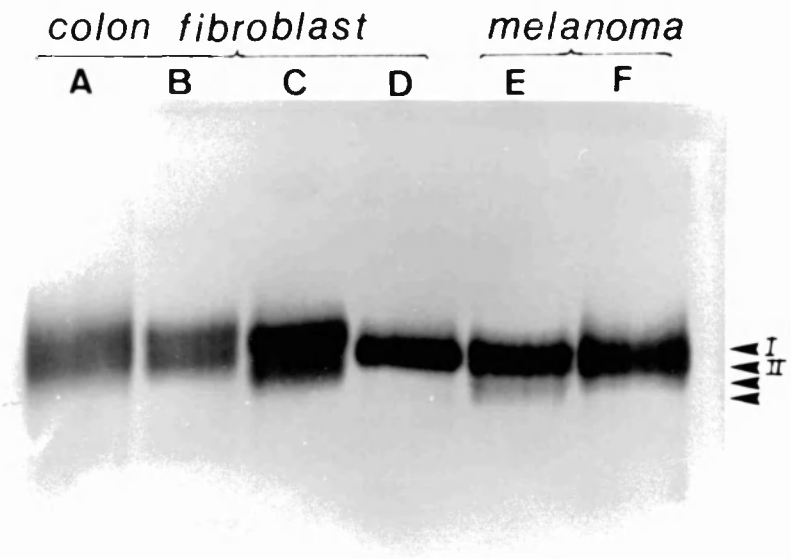


Figure 14: Immunoblots of pools B-D (fig.11d) from the lysine-sepharose fractionation of HCF tPA. *Track A contains the unfractionated HCF sample, Track E is the immunoblot of unfractionated Bowes melanoma tPA; F is unfractionated Bowes melanoma tPA from a phorbol ester stimulated cell line.*

Phorbol ester stimulated Bowes melanoma cell line (Fig.11h)

An immunoblot of unfractionated phorbol ester stimulated Bowes melanoma tPA is shown in figure 14 (track E). In contrast to the unstimulated Bowes tPA (track F), PMA stimulated cells do not produce type IV tPA. The released sugars from each of the four sets of glycoforms (pools 1-4 Fig.11h) were profiled by P4 gel permeation chromatography (data not shown). Each contained a different range of oligosaccharides, suggesting that the sets of glycoforms of t-PA bind to lysine-sepharose with different affinities which are associated differences in their glycosylation.

t-PA from β-estradiol treated MCF7 cells (Fig.11i)

The apparent absence of material in these samples (Fig.11i) was investigated by measuring the activity of unfractionated material. Interestingly, this was also very low indicating that tPA from this cell line has very low protease activity as assessed by its activity towards S2288.

Murine t-PA from C127 cell lines (Fig. 11e):

After fractionation on lysine-sepharose samples from pools 1-3 (Fig. 11e) were analysed by SDS PAGE (Fig. 15). Type D tPA is present in pools 2 and 3.

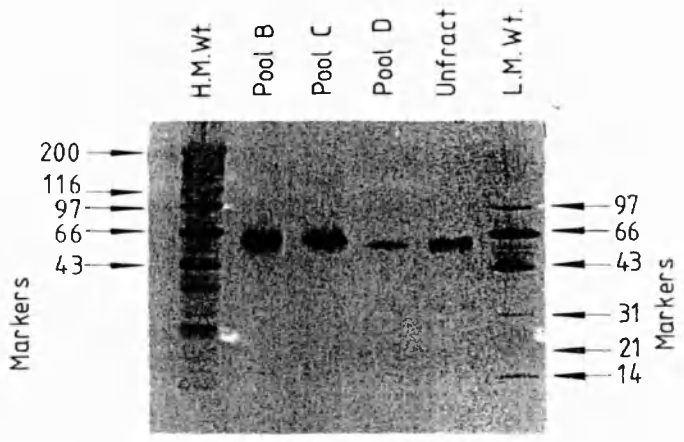


Figure 15: SDS PAGE (10% non-reducing, Coomassie stained) of Murine t-PA from C127 cells. *Tracks 1 and 6: High and low M.Wt. markers Tracks 2-4 Pools 1-3 from lysine-sepharose fractionation Track 5: unfractionated murine C127 cells tPA.*

Chinese Hamster Ovary tPA (Fig. 11g)

Fractions from the lysine-sepharose pools 2 and 3 (Fig. 11g) were analysed by SDS PAGE (Fig. 16). The overloaded gel of the unfractionated material indicates the presence of type D tPA at approximately 20kD. Lanes 2, 3 and 4 contain additional minor bands. Such bands have been noted by others (Ghislain Opdenakker, personal communication) and the amount varies in different cell lines. They appear to be the result of processing rather than glycosylation or other post translational modifications or clipping.

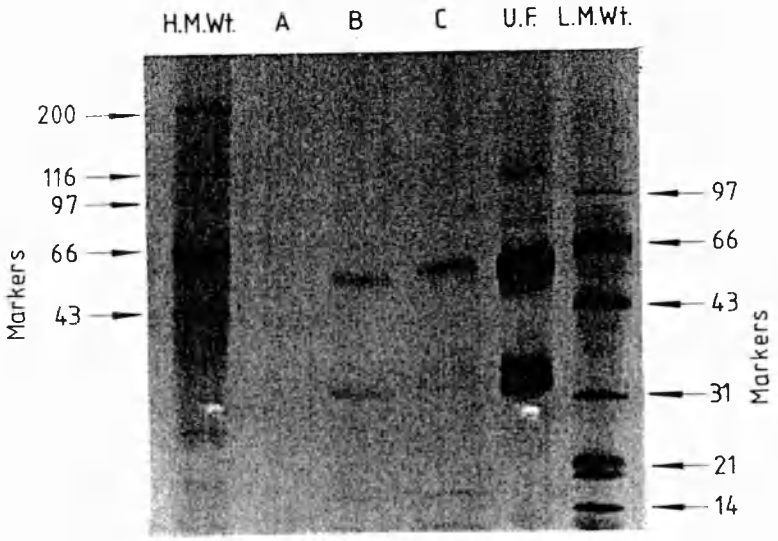


Figure 16: SDS PAGE (8-25% gradient, non-reducing, Coomassie blue stained gel) of CHO t-PA. *Tracks 1 and 5: High and low M.Wt. markers; Tracks 3 and 2 contain Type II and type I tPA respectively from the lysine sepharose column; Track 4: unfractionated CHO tPA.*

HCF t-PA from Monsanto cell lines (Fig. 11a):

The glycoform fractionation of this material on lysine-sepharose (Fig. 11a) can be compared with a similar fractionation of HCF tPA expressed in the same cell line, but in a different laboratory (Fig.11d). The data are consistent with the well established finding that cell culture conditions affect glycosylation.

BM tPA (Monsanto) and BM (Rega Institute)

The glycoform profiles of the samples fractionated on lysine-sepharose (Fig. 11b,c) show that the samples from the Rega Institute contain additional sets of glycoforms eluting between types 1 and 2 and also a larger amount of material eluting before type I.

4.4 Separation of subsets of tPA on lysine sepharose

t-PA expressed in BM, HCF, CHO and C127cell lines was fractionated by lysine-sepharose chromatography (Fig. 11c,d,g,f). Sets of glycoforms were pooled as shown and assayed for protease activity in the S2288 assay.

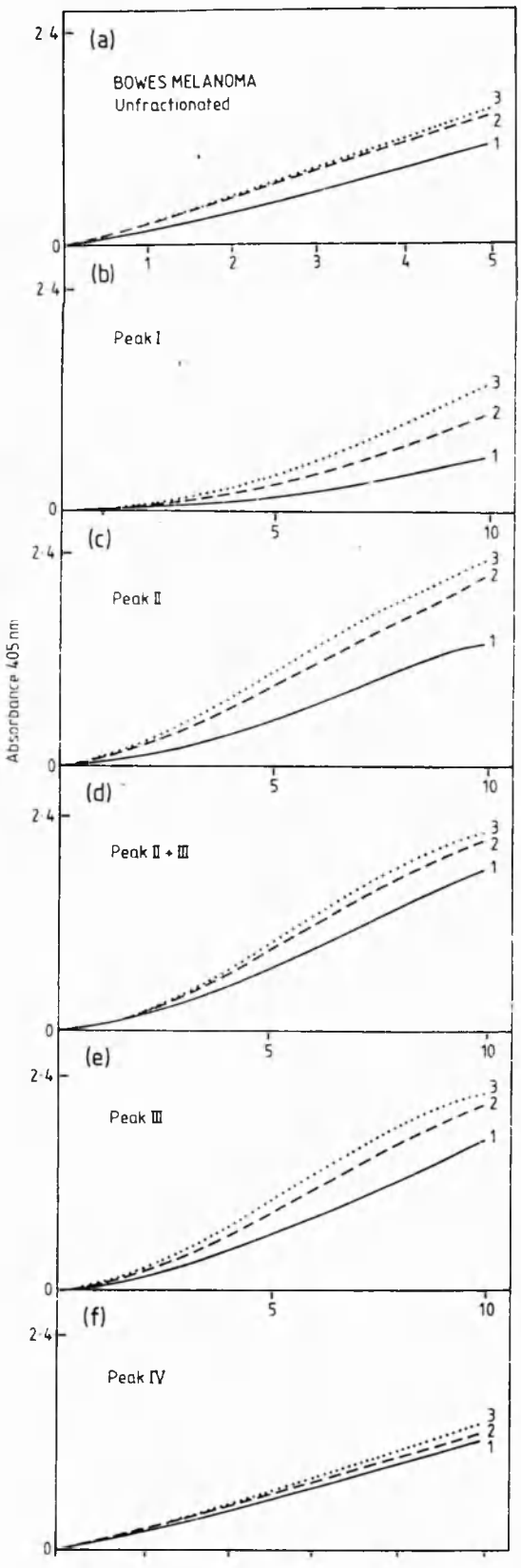
4.5 Kinetic studies of the sets of glycoforms isolated from t-PA expressed in a range of cell lines:

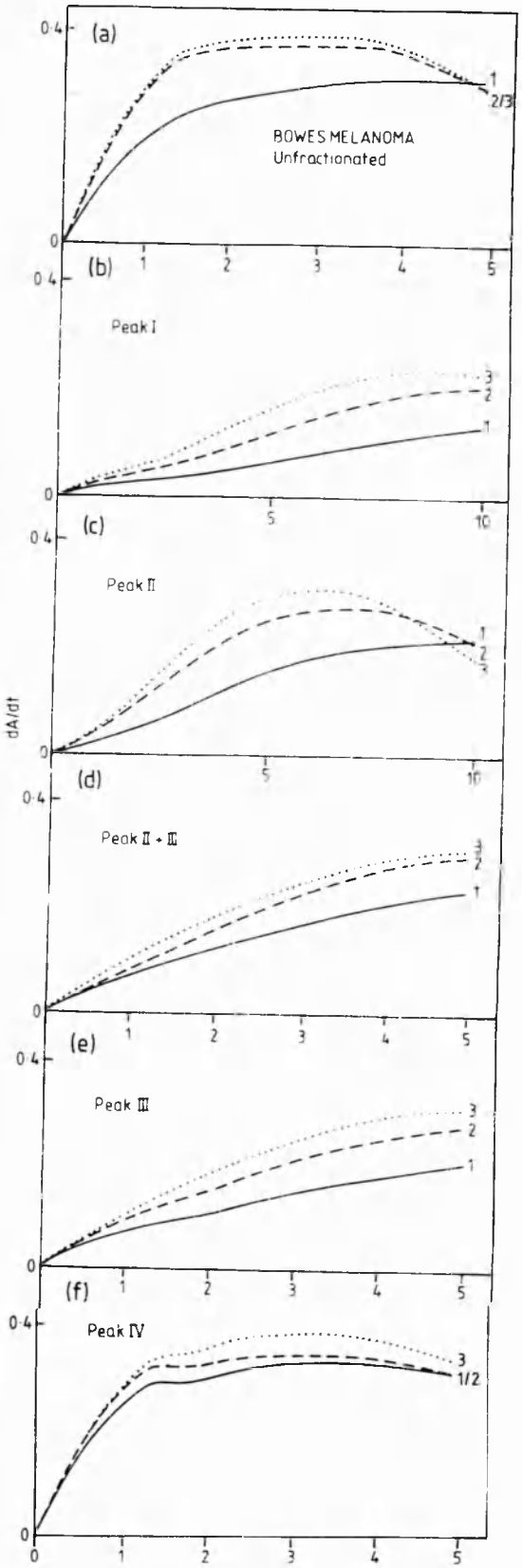
Samples with matched protease activity were withdrawn from each pool and their fibrin dependent activity compared in the indirect assay using S2251 (Fig.17a-d). The relative activities of the subsets of t-PA correlated with their elution positions on the lysine sepharose column; those fractions which bound most tightly to lysine, and therefore eluted later in the gradient, were also those with the greatest activity. In every sample of tPA tested there was an increase (4-5 times other fractions) in activity for the final pool which was eluted with approximately 0.5M arginine. SDS PAGE and/or zymography showed that this fraction contained a high molecular weight fraction (120kD) with tPA activity (e.g. Fig. 13). Table 2 gives the initial rates from the data in figure 17a-d respectively, which show the effect of adding increasing amounts of each subset of tPA glycoforms to the indirect stimulated assay (using S2251 to determine the rate of generation of plasmin).

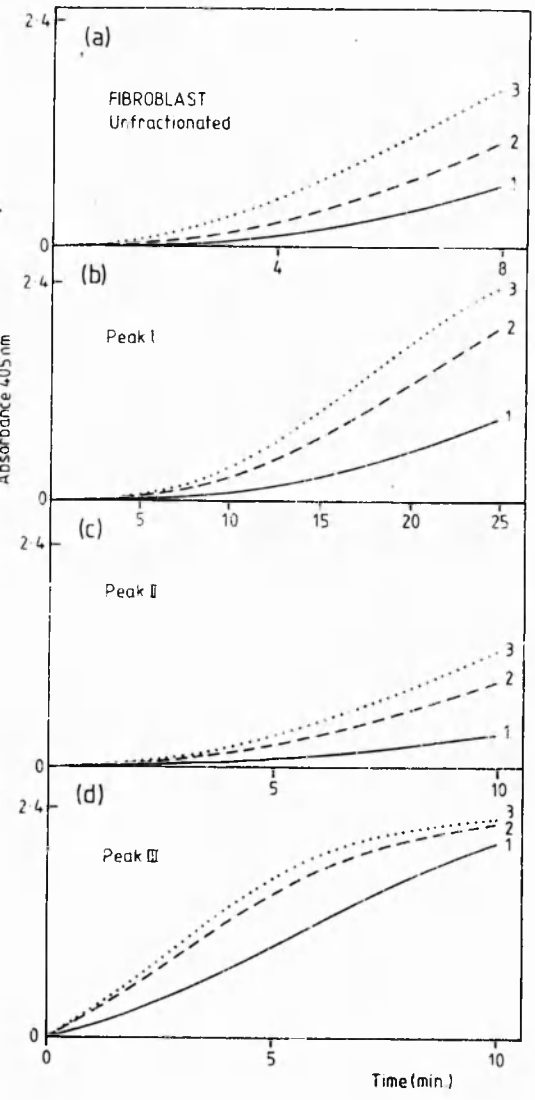
The histograms which follow (fig.18) are for C127 tPA. They show the initial rate for each type of tPA (unfractionated, Ia, I, II and D over the range of concentrations from 100-500 ng and are taken from table 1). For type I, II and D tPA there is an excess of fibrin stimulated tPA activity, the reaction is pseudo-first order and the initial rates are proportional to the concentration of tPA. Under these conditions saturation has not been reached. In the unfractionated sample and in type Ia the fibrin dependent activity is insufficient to give first order kinetics until a concentration of 300ng has been reached. The reactions at 100 and 200 ng therefore include a factor which represents the fibrin independent activation of plasminogen. The composition of fraction Ia is not known, but it is of interest because it appears to contain a subset of tPA glycoforms which have a high level of fibrin independent activity and a low affinity for fibrin.

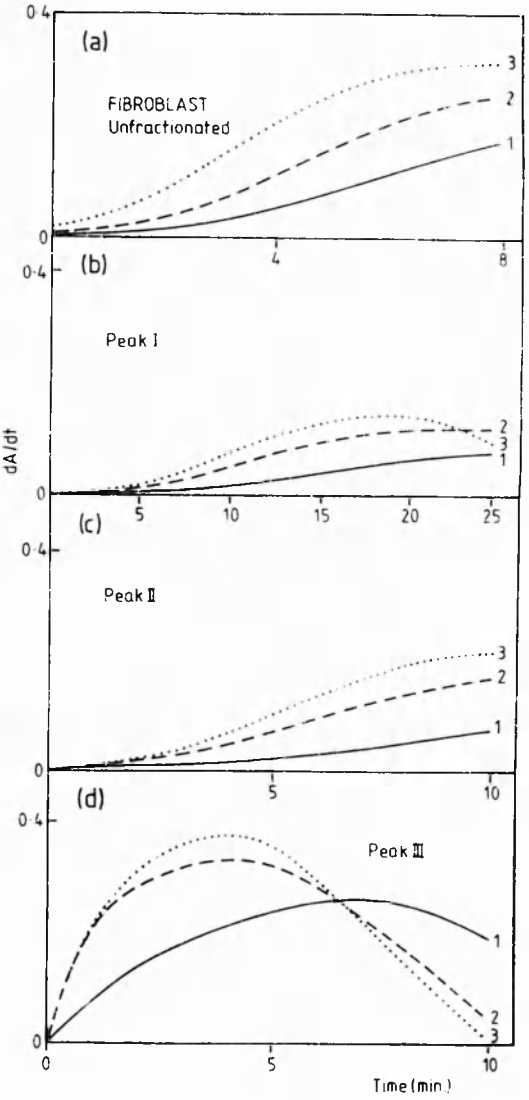
Figure 17 a-d (opposite): Fibrin dependent activities of glycoform pools of tPA from BM(a), HCF(b), CHO(c) and C127(d) cell lines.

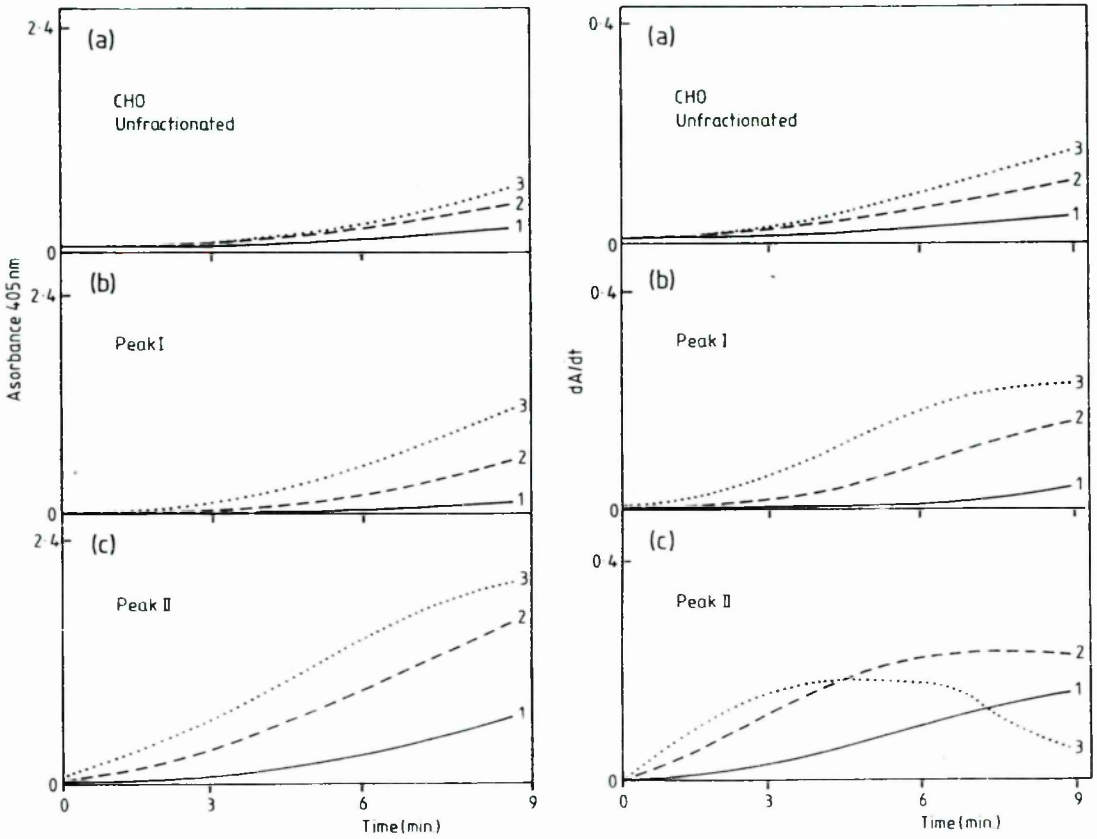
The tPA was fractionated by lysine sepharose chromatography. tPA activity in each fraction was monitored by the direct amidolytic assay using the chromogenic substrate S2288. The fractions were pooled as indicated in figure 11c,d,g and f. The activity of each pool was measured by S2288 activity and an equivalent activity of each pool was then assayed in the indirect fibrin stimulated amidolytic assay using the chromogenic substrate S2251. The lowest concentration of each pool (1) contains approximately 100ng of tPA. The activities of increasing levels (2x and 3x are also shown).

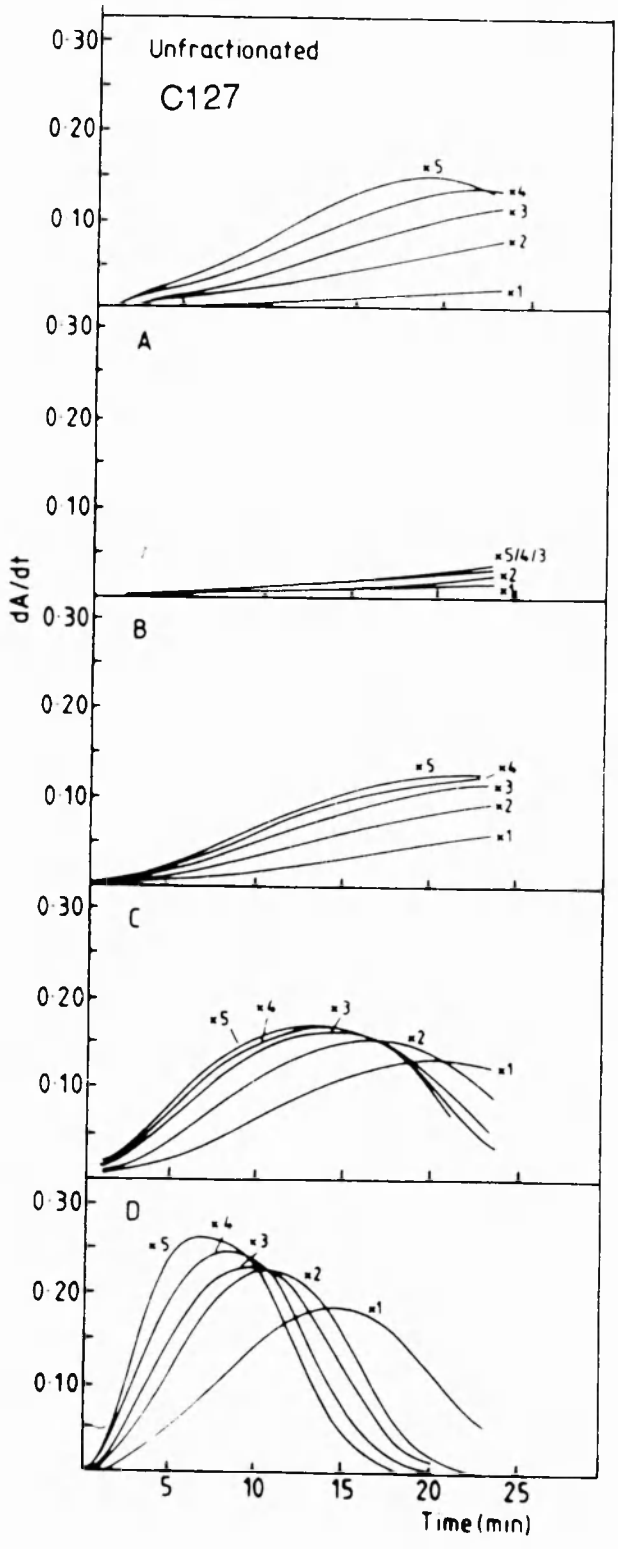

(a)
BOWES MELANOMA
Unfractionated
(b)
Peak I
(c)
Peak II
(d)
Peak II + III
(e)
Peak III
(f)
Peak IV
Absorbance 405 nm


(a)
BOWES MELANOMA
Unfractionated
(b)
Peak I
(c)
Peak II
(d)
Peak II + III
(e)
Peak III
(f)
Peak IV
dA/dt


(a)
FIBROBLAST
Unfractionated
(b)
Peak I
(c)
Peak II
(d)
Peak III
Absorbance 405 nm
Time (min.)


(a)
FIBROBLAST
Unfractionated
(b)
Peak I
(c)
Peak II
(d)
Peak III
dA/dt
Time (min.)


(a)
CHO
Unfractionated
(b)
Peak I
(c)
Peak II
Asorbance 405 nm
dA/dt
Time (min)
0
3
6
9
2·4
0·4
1
2
3

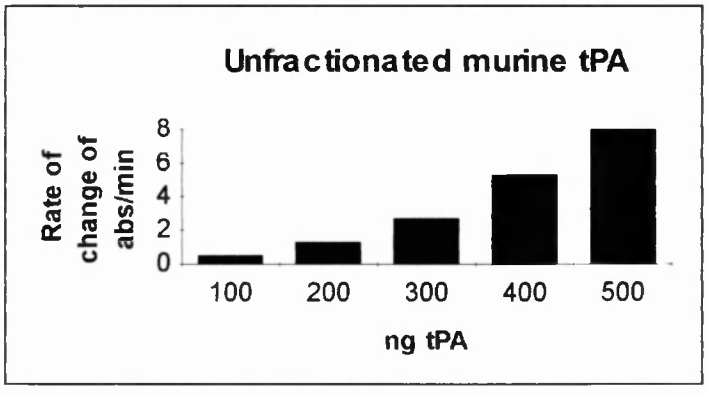

Unfractionated
C127
dA/dt
0·30
0·20
0·10
×5
×4
×3
×2
×1
A
×5/4/3
×2
B
C
D
5
10
15
20
25
Time (min)

tPA	ng	uf	Ia	I	I+II	II	D
BM	100	199	26	26	66	66	252
	200	293*	39	39	80	80	279
	300	293*	39	53	106	106	279
HCF	100	5	1.3	5		80	
	200	8	5.3	16		213	
	300	19	8	21		213	
C127	100	0.5	0.05	2.7		5	13
	200	1.3	0.13	3.2		11	31
	300	2.7	0.27	4.3		15	37
	400	5.3	0.8	5.8		19	48
	500	8	1.3	6.9		23	56
CHO	100	2.7	0.003	1.4		5	
	200	8	0.003	2.7		8	
	300	12	0.005	4.1		13	
	400	16	0.008	5.4		23	
	500	37	0.01	5.9		25	

Table 2: Initial rates for the activation of plasminogen by BM, HCF, C127 and CHO tPA

The units are ΔmOD/min. These data were obtained by taking the initial slopes from the dA/dt plots in figure 17a-d respectively. The plots show the addition of increasing amounts (from 100-300 or 500ng) of each subset of tPA glycoforms to the indirect stimulated assay using S2251 to determine the rate of generation of plasmin. The scale on the y-axes is 1cm: 13.3ΔmOD units.

* appear to be anomalous



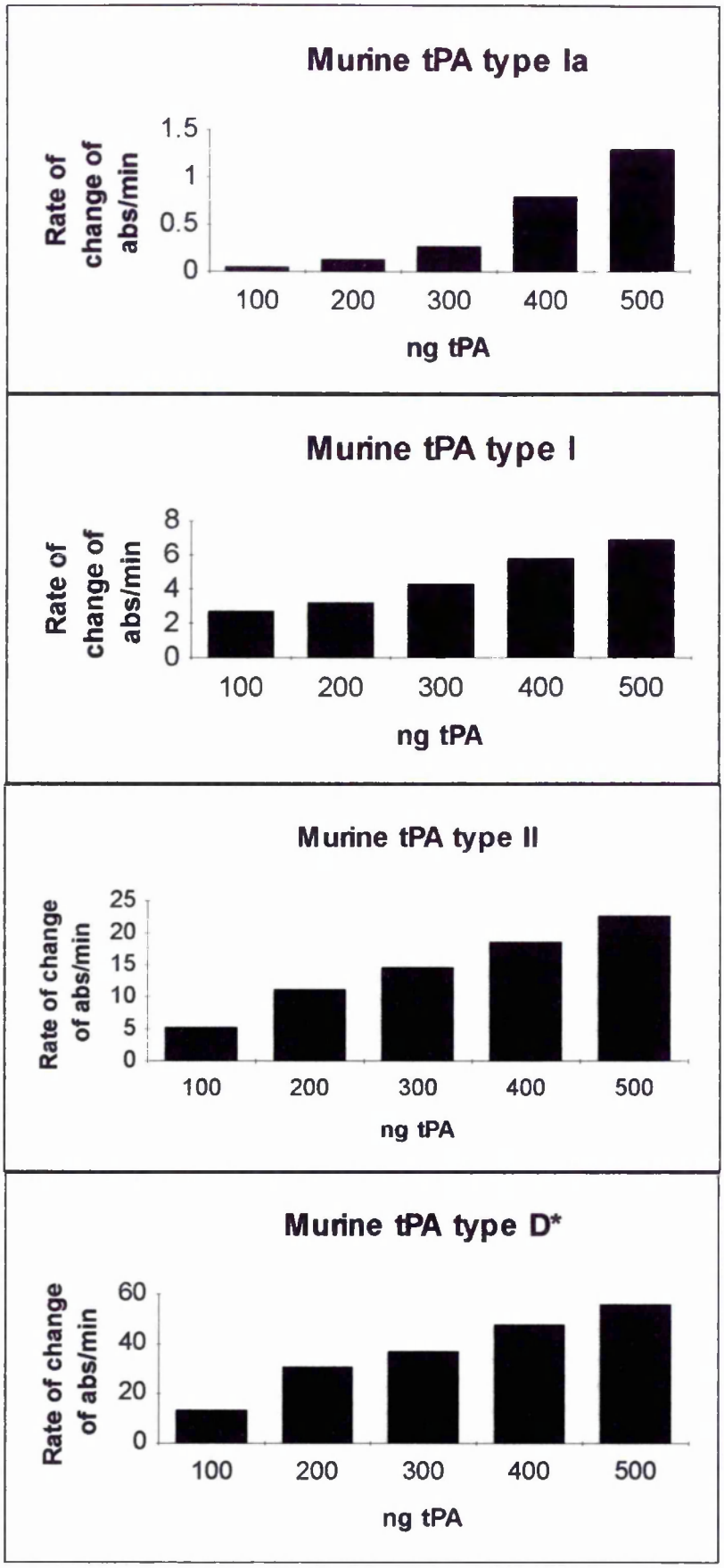


Figure 18: Histograms of initial slopes (ΔdA/dt (mOD)/ min) for peaks 1-5 from the separation of C127 tPA on lysine-sepharose plotted vs tPA concentration (0-500ng).

In Table 3 (below) the initial slopes for plasminogen activation by murine C127 and also CHO, BM and HCF recombinant tPA have been compared at a single concentration of tPA (300ng).

	C127	CHO	BM	HCF
Unfr.	2.7	12	293	19
Type Ia	2.7	0.005	39	-
Type I	4.3	4.1	53	8
Type I+II	-	-	106	-
Type II	15	13	106	21
Type D*	37	-	279	213

Table 3: Initial slopes (dA/dt (mOD)/min) for plasminogen activation by 300ng of different glycoform populations of murine C127, CHO, BM and HCF recombinant tPA. *Error +/- 10%*

Several conclusions can be drawn from the data in table 3.

(i) The activity of pool D* is cell line dependent and this may relate to the particular glycosylation of the subset of glycoforms it contains. BM D* and HCF D*tPA are the most active. The activity of type D* in BM is approximately 5x greater than type I and 2.5x greater than type II. In a similar assay using BM tPA (R) the activity of type D* in BM was found to be 3.9x greater than type I and 2.4x greater than type II (Table 3). CHO cells did not contain a fraction eluting beyond type II, but the activity of CHO type II is proportionately high compared with the relative activity of the other glycoform subsets. This suggests that in CHO cells type D tPA may have eluted with the type II fraction.

There was no lag phase in any assay involving Type D* tpA, in keeping with its high activity. Type D tPA was not recovered as a pure sample and pool D* contained on average about 50% of type II tPA. The activity of D is greater than might be suggested by the analysis of type D*.

(ii) BM tPA is more active in every subset compared with tPA from other cell lines. This is consistent with the earlier results comparing BM with HCF tPA (see Fig.8) and may reflect the smaller and less charged oligosaccharides associated with BM tPA. Glycans with a lower molecular volume and lower levels of sialic acid may allow more efficient fibrin binding and also provide less steric hindrance to the formation of type D tPA.

(iii) The pre-peak in BM, CHO and in C127 tPA has not yet been defined and may contain another interesting set of tPA glycoforms. HCF also contains a shoulder eluting before type I tPA, but this was not recovered independently. In common with peak D*, on a molar basis this peak represents a small proportion of the total protein.

(iv) Analysis of the unfractionated samples shows that the activity of BM>HCF>CHO>C127 The oligosaccharides have been profiled from each of these samples and the results show that overall BM contains the smallest, least processed sugars, followed by CHO, C127 and HCF tPA (Parekh et al 1989)

(v) The rate for unfractionated BM is anomalous, it is higher than the most active fraction. This may be experimental error or it may suggest that some further fraction of tPA with a higher activity than type D* tPA was not eluted from the column.

(vi) The reactions for pools Ia and I suggest that there is a lag phase in these fractions at low concentrations which is not present in other pools.

To examine the high activity in pool D* further and to identify the high molecular weight fraction (designated type D tPA) 500μg of BM tPA (R) was fractionated on lysine sepharose (Fig. 19). Type I, type II and fractions 85-120 (containing a mixture of type D tPA and type II tPA and designated D*) were examined further in section 3.3-3.5 in collaboration with Mr. Kazuya Mori who obtained the data. This work has been published (Mori et al 1995).

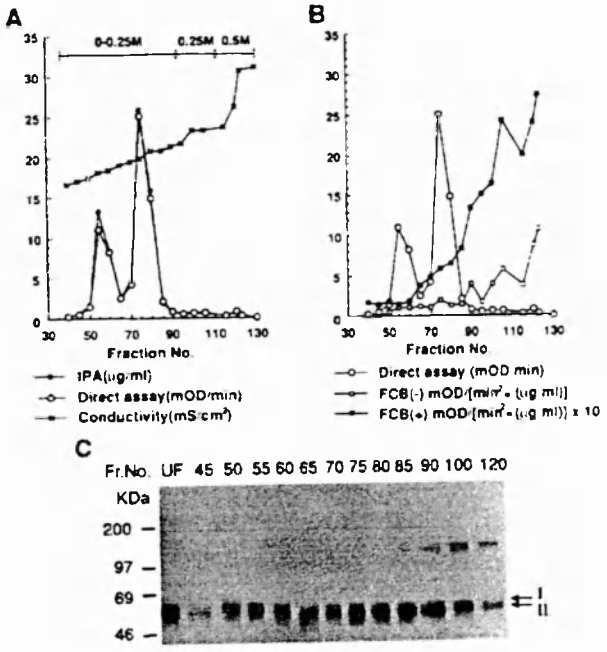


Figure 19 A,B: Separation of tPA by Lysine-sepharose chromatography and the enzyme activities of the fractions.

Bowes melanoma derived t-PA was applied to a lysine-sepharose affinity column and eluted with an L-arginine gradient (0-0.25M), followed by a wash of 0.25M L-arginine and a final step of 0.5M L-arginine. Every 5th fraction was dialysed against buffer, and both the reactivity to anti-tPA antibody and the enzyme activity were measured. ***A****, cross: conductivity of arginine in the gradient (ms/cm²); open circle: t-PA activity in the direct assay (mOD/min) using chromogenic substrate (S-2288); closed circle: antigen level of t-PA measured by polyclonal antibody based ELISA (μg/ml).* ***B****, open circle: t-PA activity in direct assay (same data as A); open squares: t-PA activity in the indirect amidolytic assay without FCB (mOD/(min²/μg/ml)); closed squares: t-PA activity in indirect amidolytic assay with FCB (mOD/(min²/μg/ml)x10). The fractions 52-63, 70-83 and 110-123 were pooled as types I, II and D* respectively.*

Figure 19C: Western blot analysis of the fractions eluted from lysine- sepharose

Each fraction was separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE, 9%), then transferred to a blotting membrane. t-PA was detected using polyclonal rabbit anti-t-PA IgG and alkaline phosphatase conjugated polyclonal anti-rabbit IgG. I and II, indicates position of type I or II tPA. The high molecular weight form of tPA (type D) is present in fractions 85-120.

4.6 Kinetic studies of Bowes Melanoma types I,II and D tPA

The activities of pooled fractions from type I tPA, type II tPA peak and type D* tPA from figure 19 were compared in three assays. The results are shown in Table 4a. In the direct amidolytic assay using S-2288, these 3 types of tPA had similar specific protease activity. In the unstimulated (without fibrin) indirect amidolytic assay, using S2251, type D* tPA showed a 5.8 times and 4.2 times higher activity than type I or type II tPA, respectively. In the presence of fibrin, type D* tPA had 3.9 times and 2.4 times higher activity than type I and type II tPA, respectively. Interestingly, the activity of type D* tPA was significantly less stimulated by fibrin than were type I and type II tPA (1:1.6:1.7 respectively).

4.7 Characterisation of type D tPA

The molecular weight (120kd) of type D tPA suggests that it may be a dimer of tPA or a complex of tPA with some other molecule of a similar molecular weight. For example, tPA can be isolated as a complex with plasminogen activator inhibitor, PAI-1. The molecular weight of this complex (approx. 110kd) is close to that of a tPA dimer. The data in figure 19B eliminates the possibility that type D tPA may be such a complex of tPA and PAI-1. In figure 20a, track 1 shows the immunoblotting by anti-tPA antibody of unfractionated tPA pre-incubated with PAI-1 and run under non reducing condition. Track 2 contains type D tPA alone run under non reducing conditions. Tracks 3 and 4 contain the same samples run under reducing conditions. Under non-reducing conditions type D tPA has a molecular weight consistent with a tPA dimer and does not migrate to the same position as either monomeric tPA (64-69kd) or tPA-PAI-1 complex. Under reducing conditions the tPA-PAI-1 complex is not cleaved and continues to run at 110kd (Track 3). Type D tPA, on the other hand, is reduced and migrates to the same position as type II tPA (Track 4). Figure 20b shows that reduction of the dimer with β-mercaptoethanol reduces the dimer to a single band migrating to the same position as type II tPA. This suggests that type D tPA is composed of type II tPA. Type D tPA was not a result of an equilibrium between monomeric and dimeric tPA since no more dimers formed in the type II fraction eluted from the lysine sepharose column even after concentration. This suggests that type D tPA may consist of a specific subset of type II glycoforms.

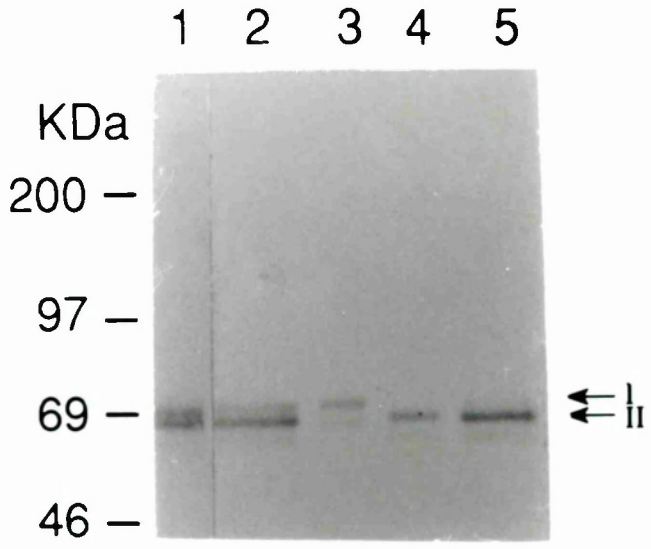


Figure 20a: Western blot analysis comparing t-PA-PAI-1 complex with type D t-PA before and after reduction.

Tracks 1 and 3; t-PA-PAI-1 complex. Tracks 2 and 4; type D t-PA. The samples were separated by SDS-PAGE (9%) and transferred to the blotting membrane. t-PA was detected by rabbit anti-t-PA IgG and alkaline phosphatase conjugated polyclonal anti-rabbit IgG.

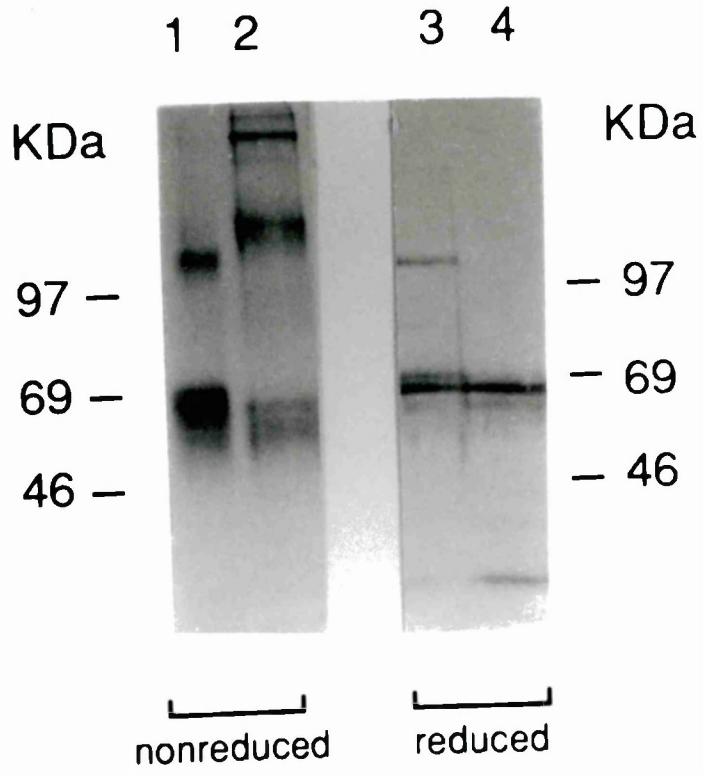


Figure 20b: Western blot analysis of types I, II and D t-PA eluted from lysine-sepharose after reduction.

All samples were reduced with 2% 2-mercaptoethanol before applying to the SDS-gel (9%). The immunodetection was carried out using rabbit anti-tPA polyclonal antibody and alkaline phosphatase conjugated anti-rabbit IgG. Track 1: mixture of type I, II and D tPA; 2: unfractionated tPA; 3: type I tPA; 4: type II tPA; 5: type D tPA.*

4.8 Effect of variable glycosylation of both Glu-plasminogen and tPA on kinetic parameters of plasminogen activation.

The rate of plasminogen activation has been shown to increase when site 184 is not glycosylated. It is also enhanced by dimerisation of the tPA. The following investigation explored whether the different glycoforms of plasminogen (type1 and type2) were equally susceptible to fibrin stimulated cleavage by type I and type II and type D tPA.

The kinetic constants for the fibrin stimulated activation of plasminogen 1 and 2 by tPA I, II and the pool containing type D tPA (D*) were determined and are in table 4b.

a. The Km values were similar for the activation of type 1 plasminogen by all the tPA variants. Likewise the Km values for the activation of type 2 plasminogen by each of the tPA variants were similar. However the Km values for interactions involving type 2 plasminogen were always lower than for type 1. This suggests that N-glycosylation at Kringle 3 in type 1 plasminogen may interfere in some way with the association of tPA and Glu-plasminogen.

b. *The kcat values* were always larger for type II tPA compared with type I tPA acting on the same plasminogen variant. This suggests that the turnover rate (kcat) is faster for the less glycosylated type II tPA variant. The Km values for the interaction of type D* tPA with plasminogen are of the same order as those obtained from the interactions of type I and type II tPA, suggesting that the association of tPA and plasminogen in the presence of fibrin is independent of the type of tPA. These data show that the Km of type II tPA with type 2 plasminogen is lower than for type I tPA with type 1 plasminogen. The other combinations of the types of tPA and plasminogen fall in between these extremes. The overall combined effect of variations in the glycosylation of tPA and of plasminogen is a 2.3-fold activity span in which the two fully glycosylated molecules have a fibrin dependent kcat/Km value which is 43% of the kcat/Km values resulting from the interaction of the two least glycosylated variants (Table 4b). A representative experiment is shown (figure 21).

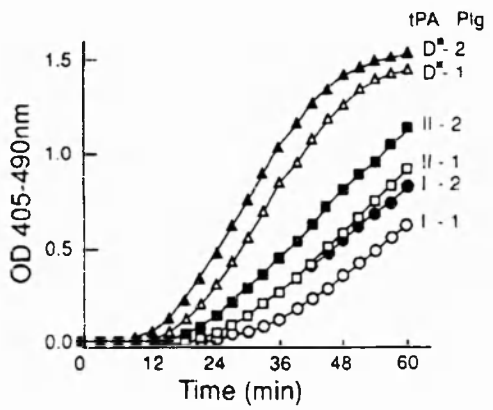


Figure 21: The kinetics of the activation of type 1 and 2 plasminogen by tPA variants I, II and D in the presence of fibrinogen fragment.

The concentrations of tPA and plasminogen in the assay were adjusted to 10ng/ml and 0.1μM respectively. Generation of plasmin was monitored by measuring the absorbance at 405-490

nm at 3 minute intervals. Open circle: type I tPA and type 1 plasminogen. Closed circle: type I tPA and type 2 plasminogen. Open square: type II tPA and type 1 plasminogen. Closed square: type II tPA and type 2 plasminogen. Open triangle: type D tpA and type 1 plasminogen. Closed triangle: type D tPA and type 2 plasminogen.

Activity of three types of tPA on three different assays

	assay(n=4)					
	direct amidolytic mOD/ [min * (μg/ml)]		unstimulated indirect amidolytic mOD/ [min^2 * (μg/ml)]		stimulated indirect amidolytic IU/ μg	
tPA	mean	SD	mean	SD	mean	SD
I	9.8	1.2	3.33	0.50	429.7	33.8
II	8.5	1.0	4.60	0.55	710.8	100.4
D *	9.0	0.8	19.29	1.75	1684.8	121.0

Table 4a:

Kinetic constants of activation of different glycoforms of plasminogen by tPA variants with fibrinogen fragment using S2251

tPA	Plasminogen	Km (nM)	kcat (sec^{-1})	kcat/Km (μM^{-1}*sec^{-1})
I	1	81.77 ± 4.72	0.118 ± 0.008	1.45
	2	47.51 ± 2.88	0.106 ± 0.014	2.22
II	1	85.63 ± 6.96	0.187 ± 0.020	2.02
	2	51.14 ± 2.27	0.170 ± 0.026	3.34
D*	1	101.23 ± 7.85	0.422 ± 0.037	4.20
	2	62.70 ± 16.25	0.388 ± 0.073	6.37

(n=4)

Table 4b:

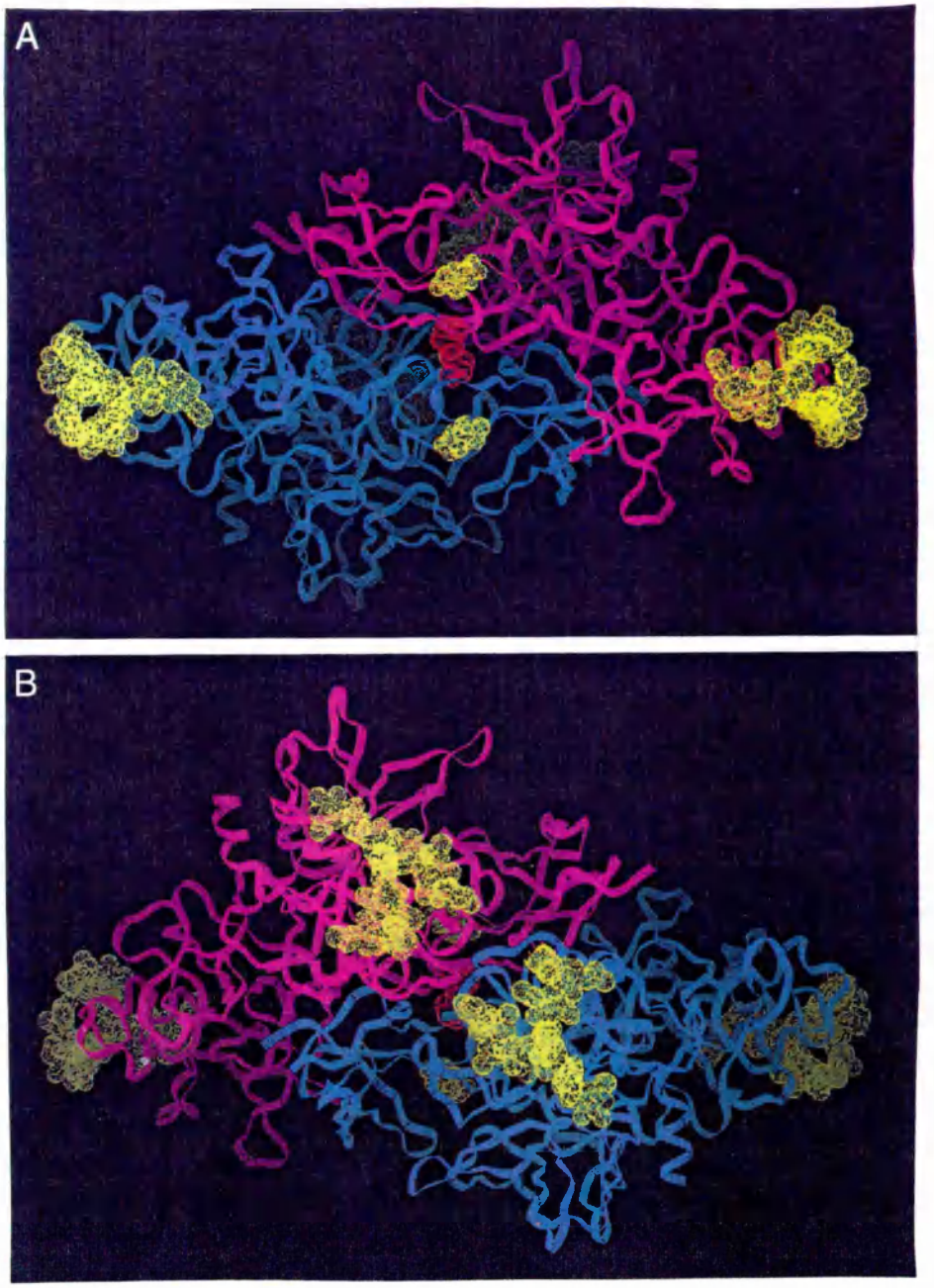


Figure 22A,B: Two views of a schematic model of a dimer of type II t-PA.

Each tPA monomer (one shown in blue, the other in magenta) is composed of five domains: a fibronectin type I finger module, an EGF-like module, two kringles and a serine protease domain. This model was constructed using the co-ordinates of the finger and growth factor pair (Smith et al in preparation) and kringle 2 (de Vos et al 1992) from human t-PA. Kringle 1 and the serine protease domains were modelled by homology. In Fig.22A the oligomannose carbohydrate at site 117 and fucose at site 61 are yellow or green. The disulphide cross-link at Cys 83 is shown in red. The complex structures at site 448 (also yellow or green) are shown in figure 22B where the molecule has been rotated through 180^{0} relative to the orientation shown in figure 22A. (Mori et al 1995).

4.9 Molecular modelling studies (in collaboration with Dr. K. A. Downing) to examine a possible structure for a tPA dimer

The data suggest that type D tPA may be composed of a particular subset of type II tPA glycoforms. It is not the result of an equilibrium involving monomers since, after the initial separation on lysine sepharose, it does not reappear in the major fraction of type II tPA even after concentration. In proposing a possible site for linking two tPA type II monomers to form the more active type D tPA there are a number of considerations to be taken into account. Firstly, type D tPA has a higher activity than other types of tPA both in the presence and absence of fibrin. This suggests that dimerisation does not interfere with the access of plasminogen to the active site (the catalytic triad, His Ser Asp) in the serine protease domain. Secondly, the dimer is dissociated into monomers by reducing agents and this suggests that a disulphide bond is involved in its formation. The primary amino acid sequence of tPA includes a number of cysteine residues which are believed to be cross linked. In addition there is a free cysteine residue at site 83 and the possibility exists that, within a subpopulation of type II tPA, this may interact with the equivalent cysteine on a second tPA molecule to form an intermolecular disulphide bond producing a homodimer.

Thirdly, the increased fibrin dependent activity of type D tPA suggests that fibrin binding sites in the finger domain, and in Kringle 2 where the lysine binding site binds degraded fibrin, (Zamarron et al 1984) remain accessible.

Fourthly, type D tPA has a similar activity to type I and type II tPA in the direct assay using the artificial plasminogen substrate, S2288 (which is a short conjugated peptide), to measure protease activity. This is consistent with the previous data which suggests that type D has a similar structure to type I and II at the active site in the serine protease domain which is the primary binding site for plasminogen. However, since type D has a 9-fold increase in activity over types I and II tPA in the unstimulated indirect assay using S2251, it is possible that dimerisation may cause a conformational change which affects a secondary binding site to plasminogen. Such a secondary site has been identified by Geppert and Binder (1992) and is in the A-chain (residues 1-275).

In collaboration with Dr. A.K. Downing a molecular model has been constructed (Fig. 22a,b) which takes these constraints into account.

5. Discussion

5.1 Patterns of glycoform distribution in tPA expressed in a range of cell lines

The type of cell has a major role in determining the extent and type of glycosylation, which is both species and tissue specific (Parekh et al 1989a). Many different enzyme reactions (typically 8 for a complex oligosaccharide such as those on tPA) are involved in the processing pathways. The profiles of the enzyme arrays, which includes their type, concentration, kinetic characteristics and compartmentalisation, reflect the phenotype of the cell in which the protein

is glycosylated. The glycosylation of recombinant glycoproteins can be exquisitely sensitive to changes in manufacturing conditions, such as the glucose concentration of the culture medium (Goochee and Monica 1990). The glycosylation pattern basically reflects the type of cell used in the expression system and the use of different cell lines can result in significant glycosylation differences. For example, there are differences in the branching structure in the complex type oligosaccharides in human colon fibroblast (hcf) tPA when compared with those of tPA derived from Bowes melanoma cell line. These suggest that Bowes melanoma cells express GlcNAc transferase (GnT) III and VII, but not GnT IV, while human colon fibroblast cells express GnT IV, but not GnT III and VII. In this chapter the glycoform populations of tPA expressed in a range of cell lines have been compared by eluting the glycoprotein from a lysine-sepharose affinity column with an arginine gradient to partially resolve the subsets of glycoforms. The profiles were obtained by analysing the eluted fractions specifically for protease activity using an artificial substrate S2288. This is a relatively rapid screening method which does not require the tPA sample to be highly purified and may therefore be useful for following glycosylation changes during cell culture. This approach also has an advantage in that the glycosylation is profiled directly at the protein level allowing comparisons to be made of the glycoform patterns in different samples. In contrast, the more commonly used method for profiling glycosylation variations of tPA in cell culture systems involves prior purification of the tPA followed by gel filtration of the released and radiolabelled glycans. This method has the intrinsic disadvantage which is that, for a protein with more than one glycosylation site, the analysis of the released glycan pool cannot provide information about the particular combinations of glycans which are attached to the protein nor can it portray the complexity and distribution of the glycoform populations.

The individual profiles obtained in this study (Fig. 11) were characteristic of the cell line and culture conditions in which the tPA was expressed, and, predictably, there were many differences in the glycosylation patterns. tPA from unstimulated BMR (BM tPA prepared in the Rega Institute(R)), HCF from serum positive and serum free media (R), Phorbol Ester treated BM(R), β-estradiol treated MCF-7(R), C127(Monsanto (M)), CHO(M), HCF(M) and BM(M), and in, a separate experiment, BM (Rega) were all applied to a lysine sepharose column. Sets of glycoform populations were successively eluted from the column with an arginine gradient (0-0.25M).

A comparison of the glycoform profiles of two HCF preparations obtained from serum containing media (Fig. 11a and d) shows that there are variations between the batches. This is consistent with the well established fact that differences in culture conditions affect the glycosylation process (Goochee et al). A direct correlation between protease activity in the direct unstimulated assay using S2288 and protein concentration, as judged by α-tPA antibody, has been established for BM tPA (Mori et al 1995). Therefore it is of interest to note that the profile of tPA from β-estradiol treated MCF7 cells (Fig.11h), obtained by measuring the S2288 protease activity of each sample, suggests that there is no active tPA in the fractions. In fact the activity of the unfractionated sample towards the peptide substrate, S2288, which contains

only three amino acid residues and binds in the active site was also very low. The reason for the low protease activity of this particular sample of tPA towards this peptide is not known, however it seems unlikely to be related to glycosylation. Examination of the molecular model (Fig. 22) indicates that even the largest tetra-antennary glycan so far identified on tPA would be unlikely to hinder sterically such a small peptide from entering the active site.

On lysine-sepharose columns glycoforms with the least affinity for lysine eluted first. All of the tPA samples contained two major peaks containing type I tPA, with three occupied glycosylation sites, eluting first and type II tPA, with two occupied sites, eluting later. However, the relative proportions of these glycoforms varied from 60% of type II in HCF(Monsanto) to 30% in C127 tPA. In addition to the two major components all of the samples contained tPA eluting at positions before, between or following the two main peaks. This indicates that both type I and type II tPA consist of a wide range of glycoforms with different affinities for lysine and, by implication, for fibrin. Together with the gel of BMR in figure 13b, which shows the presence of type III and IV tPA, the lysine sepharose profiles and activity studies suggest that most tPA preparations are more heterogeneous than has been suggested previously. In further support of this evidence for heterogeneity, a previously unreported dimer of tPA type II (type D) with strong affinity for lysine was identified and characterised in this chapter. It is possible, but not so far demonstrated, that the additional peaks may also contain glycoforms of tPA which are unglycosylated or display variable site occupancy at sites other than Asn 184. Interestingly, C127 Type I tPA, expressed in the presence of dimethylmannoside (dMM) with oligomannose structures at 184, bound lysine with an affinity between type I and type II (Howard et al 1991). This suggests Type III and IV tPA identified in the BM cell line, which contains an increased proportion of oligomannose structures, may consist of Type I tPA with oligomannose structures at 184. Howard et al 1991 found that tPA from type I with either charged or neutral complex sugars had the lowest affinity for lysine. This suggests that the subset of glycoforms, type Ia, eluting before type I may contain sugars of this type.

5.2 Some factors which influence fibrinolytic activity

5.2.1 Affinity of tPA glycoforms for lysine

In every sample of tPA the fibrin stimulated activity of the glycoforms increased with the affinity of the subsets for lysine sepharose (Fig. 17, table 1). In figure 11, which shows the lysine sepharose elution profiles of the tPA samples, minor variations in the gradient mixing may have occurred in the different preparations. This has been accounted for by measuring the conductivity of every fraction eluted from the column. In figure 11 the profiles are therefore aligned according to the arginine gradient. It can be seen that, in general, the peaks in the BM samples (11b,c,h,j) elute at a higher concentration of arginine than HCF tPA (a,d). This indicates that overall the glycoforms of BM have a stronger affinity for lysine than HCF tPA, and suggests that BM tPA will be more active in the fibrin stimulated assay. This is consistent with the kinetic analyses of unfractionated samples which showed that the activity of BM (M) is

2-3 times greater than HCF (M) tPA (Fig.7 and Table 1). Kinetic analysis of a number of unfractionated tPA samples from the Rega Institute disclosed a range of fibrin stimulated activities in which BM>HCF>CHO>C127. Some of the possible reasons for these differences in activity are discussed below:

5.2.2 Relative proportions of type Ia,I,II and D

In the first instance activity differences between samples are related to the relative proportions of type Ia, type I, type II and type D tPA, which bind to lysine with increasing affinity. In this respect the proportion of type D tPA is particularly significant.

In the C127, BMR and HCF cell lines a proportion of tPA (approximately 1-4%) was expressed as a glycoform with a molecular weight of approximately 120kD. Although this fraction represented a small proportion of the total protein it had a dramatically increased activity. The material was characterised as a dimer of type II tPA (type D tPA), and may be a product of cell culture preparations and therefore not physiologically relevant. However in view of the widespread therapeutic use of recombinant tPA it may be important to recognise this species, which has greatly increased activity both in the fibrin independent (Table 4a (BM)) and fibrin stimulated reactions (Table 4a (BM)). The tightly bound fraction, designated D*, which in each case contained type D diluted with type II tPA, had from 2-10x the activity of type II tPA in the fibrin dependent assay depending on the cell line (Table 2 (CHO, C127, HCF and BMR) and Table 4a,b (BM)). The activity of pure type D tPA is therefore in excess of these values. The effect of dimerisation on clearance has not been studied, however inspection of the proposed molecular model (Fig. 22) suggests that access to the O-linked fucose residues at Thr61, which has been reported to bind to a hepatic receptor (Hajar et al 1993), would not be significantly hindered. This suggests that type D tPA is composed of type II tPA. Type D tPA was not the result of an equilibrium between monomeric and dimeric tPA since no more dimers formed in the type II fraction eluted from the lysine sepharose column even after concentration. This suggests that type D tPA may consist of a specific subset of type II glycoforms. The data in figure 12 also suggest that type D tPA, which throughout all the cell lines elutes in a relatively narrow conductivity band (0.76ms/cm), may consist of a specific subset of type II glycoforms. The amount of type D tPA in each sample was qualitatively assessed by gels to be C127>BM>HCF>>>CHO over an estimated range of 1-4% of the total protein. To assess the influence of increasing the percentage of type D (using the data in table 4a), the increase in overall activity of BM tPA containing 4% type D compared with 1% type D is 11%. However, in itself this contribution to the overall activity, although significant, is not sufficient to explain fully the observed differences in rate. For example, C127 tPA contains a higher proportion of type D than any of the other tPA samples, but its activity is lower. C127 also contains a relatively lower proportion of type II:type I tPA compared with other samples. This may suggest that a significant proportion of type II tPA from C127 consists of a subset of glycoforms which is able to form dimers through intermolecular disulphide bonds. In figure 19C (track 2), which is an immunoblot of a non-reducing gel of pool D*, there is evidence of a band with a molecular

weight of approximately 240kD. This suggests that type II tPA may form tetramers or possibly even larger aggregates.

5.2.3 Fibrin independent activity

Another interesting feature which emerges from the kinetics data in Table 2 and figure 17 is that the subset of glycoforms type Ia have a high level of fibrin independent activity, a low affinity for fibrin, and a low fibrin dependent activity. The amount varies in each cell line and culture, the PMA stimulated BM cell line contains about 18%, while the HCF (M) contains less than 5%. The glycosylation of fraction Ia is not known, but the discussion which follows (section 5.2.4) suggests that it may contain type I tPA with large complex glycans at site 184. In figure 12 the conductivity range of type I tPA from different samples of tPA was narrow (1.4ms/cm) suggesting that it might consist of a specific subset of type I glycoforms. This possibility is consistent with data from Howard et al (1991) which found that tPA from type I with either charged or neutral complex sugars had the lowest affinity for lysine. The contribution of this set of glycoforms and of unstimulated activation of plasminogen to the overall activity is not known. Increased levels of glycoforms which activate plasminogen in the absence of fibrin may increase the fibrin stimulated activity of tPA by degrading fibrin to produce terminal lysine residues. The ratio of fibrin independent activity of type I + Ia:II:D in BM was 1:1.4:5.8 (Table 4a). The corresponding ratios for fibrin stimulated activity are 1:1.6:3.9. The extent to which the fibrin dependent activity of different types of tPA depends on fibrin independent activation, perhaps by particular subsets of glycoforms, is not known.

5.2.4 The detailed glycosylation at site 184

The affinity of type I tPA for fibrin and thus the fibrin dependent activity of type I tPA and the overall activity of a sample, may also be influenced by the specific set of sugars at site 184. These are in a position to obstruct sterically the lysine binding site in K2 and examination of the molecular model (Fig.22) indicates that longer, more highly branched sugars would be more effective in fulfilling this role. Site analysis of HCF and BM (Parekh et al 1989) shows that the major populations of asialo glycans at 184 are smaller in BM (22.5% complex (C):16gu; 11% (C):15gu and 10.1% oligomannose:14gu) compared with HCF (39.3% (C):15gu; 12.4% (C):16 gu and 11.1% (C):18gu), although the remaining populations of glycoforms (56.4% BM; 37.2%HCF) are in the range of 9-14 gu in BM compared with 14-22gu in HCF. Overall Bowes has a higher percentage (35.2% (M8-M4) compared with 13.8% (M7-M4) in HCF) of oligomannose glycans at site 184. A study of the mannose 9 (M9) glycan (Woods et al 1994) indicated that there is tight hydrophobic stacking of each of the three terminal residues with the adjacent penultimate residue in the chain. This suggests that the M7-M8 oligomannose structures in tPA may not cover significantly more of the protein surface than a M6 glycan. These data suggest that a contribution to the increased rate observed for BM compared with

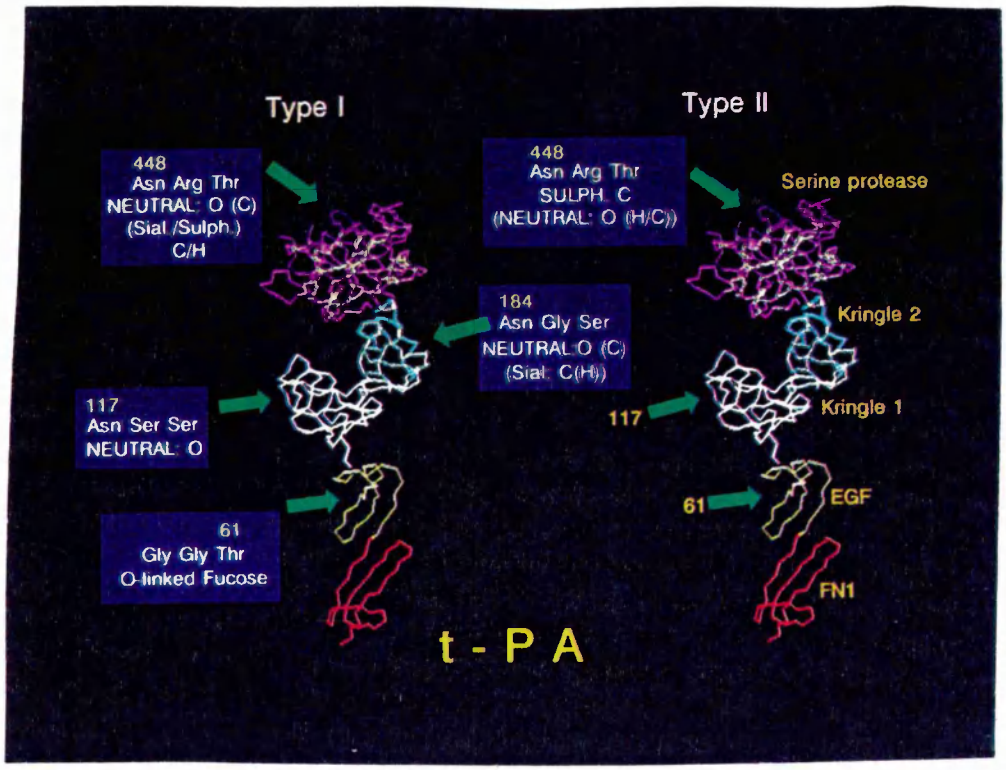


Figure 23: Glycosylation at site 184 affects the glycan composition at site 448

The classes of glycan structures present at Asn 448 on tPA depend on the site occupancy of Asn 184 while those at site 117 do not. In type I tPA the major species are neutral and of the complex or oligomannose type, and only 72% of the glycans are sialylated or sulphated complex or hybrid sugars (Wing et al - unpublished data). In type II 13% of the glycans are complex or sulphated structures. The sugars at site 117 are almost invariably oligomannose in both type I and type II tPA regardless of the cell line in which the tPA is expressed.

HCF may come from specific differences in the glycosylation of site 184. This is consistent with the finding by Howard et al (1991) that oligomannose structures at site 184 in C127 tPA increased the activity of tPA 2-3 times, however removal of sialic acid from tPA had no effect. Binding to lysine and fibrin is mediated by the K2 domain, and while the finger region also binds to fibrin, it does not interact through lysine (van Zonneveld et al 1986). Alteration of the glycosylation at site 448 to oligomannose structures did not affect lysine binding (Howard et al 1991).

5.2.5 Overall glycosylation and charge profiles

Oligosaccharide profiling of the total asialo glycan pool (Parekh et al 1989) indicated that overall BM contained the smallest, least processed sugars, followed by CHO and C127 with HCF tPA containing the largest, most processed glycans. There was, therefore, no direct correlation between the overall glycosylation pattern and activity. All of the tPA samples were shown to have identical amino acid sequences, therefore these data suggest that the rate differences may be related to specific variations in the glycosylation pattern at particular sites.

The charge profiles of the four samples indicate that BM tPA contains the highest proportion of neutral glycans, followed by HCF, CHO and C127 and interestingly, this order correlates with the relative activities of the samples. Site 117 is glycosylated almost entirely with neutral oligomannose structures therefore charge differences relate to sites 184 and 448 in type II tPA and only to site 448 in type II. It is therefore interesting to note that in BM type II tPA there are no sialylated structures at site 448, and all the charge is due to sulphate (Wing et al unpublished data). In type I the charge distribution at 448 includes a smaller proportion of charged sugars than in type II but most of them are sialylated (Fig. 23). This suggests that increased sialylation at site 448 in type II tPA or at either site 448 and/or site 184 in type I tPA may reduce the activity of tPA. In support of this, Howard et al (1991) also found that desialylation or the presence of oligomannose structures at 184 or 448 increased activity. However, neuraminidase treatment did not increase lysine binding (Howard et al 1991), suggesting that the lower activity of tPA containing increased levels of sialylated structures noted in this chapter is not related to the affinity of this set of glycoforms for lysine, but may modulate some other aspect of plasminogen activation. The data from this chapter also suggest that the substitution of sulphate for sialic acid may increase the enzyme activity since, in contrast to BM tPA, the less active HCF, CHO, and C127 derived enzyme contained charge due only to sialic acid (Parekh et al 1989). BM cells are derived from a neural cell line, and interestingly studies of other neural glycoproteins have noted the presence of high levels of sulphated compared to sialylated glycans. The function of tPA in the brain is not associated with fibrinolysis, but is related to subtle modulations of cell/cell and cell/ECM interactions which are required for neuronal and synaptic plasticity (Sappino et al 1993).

The factors which combine to give increased fibrinolytic activity to tPA may therefore include: (i) an increase in the proportion of type II tPA (monomers and dimers) which reflects a decrease in the efficiency with which site 184 is modified by the $GlcNAc_2Man_9$ oligosaccharide precursor (ii) an increase in the proportion of type II glycoforms which dimerise through intermolecular disulphide bonds (iii) a decrease in sialylated structures and an increase in either neutral or sulphated glycans at site 448 (iv) a decrease in sialylated and fully processed glycans at site 184 in type I. (v) The level of fibrin independent activation is also a significant parameter. Interestingly, high affinity binding to fibrin may not invariably be the dominant factor, for example, single chain tPA has greater affinity for fibrin but it is 20-50 times less active than the 2-chain form (Wittwer and Howard 1990; Tate et al 1986).

It is interesting to speculate how significant variations in the levels of type I and type II tPA might arise. The proportions of type I:type II are approximately 50% of each in the human cell lines studied in this chapter and in others (personal communication Dr. M. Spellman). However different proportions have been noted in some cell lines, such as murine tPA in this chapter where the percentage of type I:II is about 70:30 and in uterine tPA where there is an increased level of type I tPA and a ten-fold decrease in activity compared with tPA from porcine heart muscle (Rijken et al 1979). This suggests that under some circumstances, the percentage of occupied sites might vary in human tPA. Such a situation might develop in endothelial cells close to the site of an intravascular fibrin clot. If unusually high levels of tPA are being expressed the addition of the $Man_9GlcNAc_2$ precursor to Asn 184 may become less efficient and this would result in a reduction in the number of glycosylation sites being filled. The consequence of this may be an increase in type II tPA and the release of a form with more fibrin dependent activity at the site of the clot. Moreover, increased levels of tPA expression at such sites may in itself lead to inefficient processing of other glycans and to lower levels of sialylation, if the proportion of glycoproteins to be processed rises to levels at which the concentrations of the glycosylating enzymes are insufficient to process every glycan fully. This may also lead to the release of a more active form of tPA. Interestingly, tPA expressed in CHO cells cultured in medium containing dithiothreitol, which prevents disulphide bond formation, were fully glycosylated at site 184. The tPA was normally folded and secreted when the reducing agent was removed and only type I tPA was identified (Allen et al 1995). This suggests that the rate of folding and disulphide bond formation of tPA determines the extent of occupancy of site 184.

5.3 tPA glycosylation affects the turnover rate of the enzyme while the glycosylation of plasminogen affects the association of the ternary complex.

Both the fibrin dependent and independent activity of BM t-PA was significantly higher than that of HCF t-PA (Fig.8, 9). In fibrin stimulated plasminogen activation the Km values for both types of t-PA were similar, however the Vmax value for BM (0.2 ΔOD/min) was over twice the Vmax for HCF (0.095 ΔOD/min). This indicates that the glycosylation of tPA does not affect the formation of the ternary complexes, but that it does modulate the turn over rate. Some insight

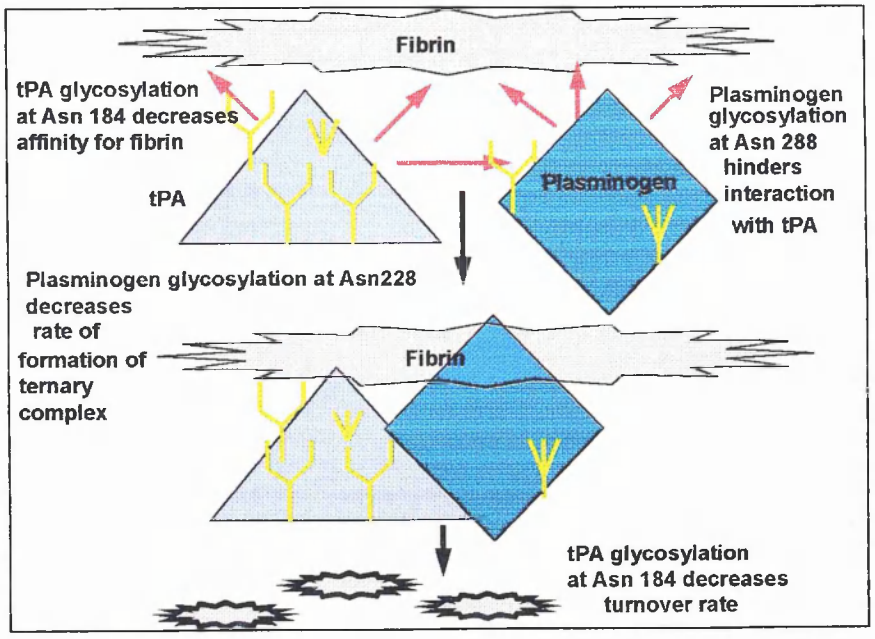


Figure 24: The effect of variable site occupancy on the Km and kcat values for the fibrin dependent activation of plasminogen by tPA

This schematic drawing is intended to convey the main conclusion from this chapter which is that variable glycosylation of tPA at Asn 184 is a critical factor in the mechanism which controls the fibrinolytic activity of tPA, affecting the turn over rate (kcat) but not the formation of the ternary complex (Km). The factors which combine to give increased fibrinolytic activity to tPA are: (i) an increase in the proportion of type II tPA (ii) an increase in the proportion of type II glycoforms which dimerise through intermolecular disulphide bonds (iii) a lower proportion of sialylated structures and an increase in either neutral or sulphated glycans at site 448 in either type I and /or type II (iv) a lower proportion of sialylated structures at site 184 in type I. kcat and Km values are shown in table 4.

reducing agent was removed and only type I tPA was identified (Allen et al 1995). This suggests that the rate of folding and disulphide bond formation of tPA determines the extent of occupancy of site 184.

5.3 tPA glycosylation affects the turnover rate of the enzyme while the glycosylation of plasminogen affects the association of the ternary complex.

Both the fibrin dependent and independent activity of BM t-PA was significantly higher than that of HCF t-PA (Fig.8, 9). In fibrin stimulated plasminogen activation the Km values for both types of t-PA were similar, however the Vmax value for BM (0.2 ΔOD/min) was over twice the Vmax for HCF (0.095 ΔOD/min). This indicates that the glycosylation of tPA does not affect the formation of the ternary complexes, but that it does modulate the turnover rate. Some insight into the molecular details of this finding come from a study by Horrevoets et al (1993) who demonstrated that deletion of either the finger domain or K2 on tPA results in an increase in Kd for the binding of tPA to fibrin. This suggests that binding of tPA to fibrin, mediated simultaneously by both the finger and the K2 domains, is essential for a correct orientation of the enzyme on the fibrin polymer to give the optimal stimulation by fibrin on plasminogen activation. The data from this chapter suggest that the interaction between tPA and fibrin is critical, not for the formation of the ternary complex, but for the turnover rate which involves the re-arrangement of the complex, the cleavage of plasminogen and the release of tPA from the cleavage product, plasmin. This is consistent with a study by Paoni et al (1993) in which mutations of tPA which increased its specificity for fibrin caused a reduction in kcat, but not in Km. Examination of the molecular model (Fig. 4b) shows that glycosylation at Asn 184 shields a binding site for lysine in kringle 2, suggesting that variable sequon occupancy confers on the protein a means to modulate fibrin binding and as a result the turnover rate of tPA, which is a critical event in the fibrinolysis pathway. Consistent with this, data from this chapter (Table 2), which was discussed earlier, suggested that the differences in activity between the different types of tPA may be a result of glycosylation differences at one particular site, rather than an overall change in glycosylation.

To confirm the finding that tPA glycosylation influences the turnover rate and to explore the factors which might govern the formation of the ternary complex, type 1 and type 2 plasminogen were interacted with type I and type II BM tPA in the fibrin stimulated assay. Fig. 21 and Table 4b show that type 1 and type 2 plasminogen exhibit different kinetics when interacting with type I or type II tPA in the ternary complex with fibrin. The variable glycosylation of both molecules affects the overall rates of the reactions (kcat/Km) which range from $1.5\mu M^{-1}sec^{-1}$ for interaction of the two fully glycosylated enzymes to $3.3\mu M^{-1}sec^{-1}$ for the least glycosylated ones (Table 4b). The variable glycosylation of tPA affects kcat, the absence of glycosylation at Asn289 on plasminogen type 2 is associated with a decrease in Km for activation by both types of tPA. These data indicate that the glycosylation status of tPA affects only the turnover rate and not the formation of the ternary complex and are consistent with the

N-glycosylation of intact Glu-plasminogen had no effect on fibrin binding. However the glycosylation decreases lysine binding affinity, suggesting that, after cleavage of fibrin by plasmin, which exposes new plasminogen binding sites (Suenson et al 1984), the N-linked glycan may decrease the binding to terminal lysine. Interestingly, glycosylation at Asn289 does not significantly affect the binding constants of ligands for the two allosteric sites (K4 and K5), nor the forward rate of α– to β- conformational change. It does, however, reduce the rate of the reverse β- to α- conformational change which is from an open conformation to a compact form, and this may be important in the formation and rearrangement of the ternary complex with fibrin and tPA (van Zonneveld et al 1986)

5.4.2 O-glycosylation at Thr 345 (figure 3)

Type 2 plasminogen, which contains only O-linked glycans, exists in six glycoforms which are distinguished from each other by the amount of sialic acid (from 1.3-13.65 mol/mol of protein) they contain. All forms were converted to plasmin by tPA, however the rate of the fibrin dependent conversion of plasminogen to plasmin increased as the proportion of sialic acid decreased ($3.77x10^4$ - $1.65x10^5$ $M^{-1}s^{-1}$) (Pirie-Shepperd et al 1995). This suggests that, in addition to site occupancy, microheterogeneity also plays a role in the regulation of fibrinolysis. Differently sialylated plasminogen may represent systemic pools of this proenzyme that can be activated at different rates on the fibrin clot, ensuring persistence of fibrinolytic activity.

5.5 The interaction of Type D tPA with plasminogen and fibrin

In a direct amidolytic assay, using the chromogenic substrate S-2288 to probe the interaction of a small peptide with the active site in the tPA serine protease domain, all three types of tPA showed a similar specific activity (activity/monomer) (Table 4a). This suggests that the proteolytic cleavage site is unaffected by either glycosylation or dimerisation. In contrast, in the indirect unstimulated amidolytic assay, which probed the interaction of tPA variants with native unfractionated human Glu-plasminogen substrate, activities were in the ratio 4:1:1.2 for type D* tPA : type I : type II tPA. This suggests that dimerisation of tPA results in an increased rate of plasminogen activation in the absence of fibrin. The interactions of type 1 and type 2 plasminogen with types I, II and D* (type II + type D) tPA have kcat/Km values which range from 1.5 to 6.4. The data in Fig. 21 shows that the rate of production of plasmin is greatest when the least glycosylated variants of both tPA and plasminogen are involved and the tPA type II is in the dimeric form (type D). The Km values for the interaction of type D* tPA with plasminogen were of the same order as those for the interactions of type I and type II tPA, consistent with previous data which suggested that the association of tPA and plasminogen in the presence of fibrin is independent of the type of tPA, and showing that the type II dimer reacts in a kinetically similar manner to the monomer in this step. However, the kcat values for type D* tPA with type 1 and type 2 plasminogen are significantly greater than for either type I or type II tPA. This suggests that the pure dimers of type II tPA (type D) have an increased turnover rate. Since the rate constants are for type D diluted with type II; the absolute value for

D will be even higher. The kcat/ Km value for the interaction of type D* tPA with type 2 plasminogen is 4 times as large as the value for type I tPA and type 1 plasminogen. The fibrin dependent catalytic activity of tPA therefore spans a 4 fold range and is dependent on both the glycosylation of tPA and plasminogen and the dimers of type II tPA. The Km values, on the other hand, are modulated only by glycosylation at site 289 on kringle 3 of plasminogen.

Type II and type D tPA bind more tightly to fibrin than type I and are therefore more likely to remain attached to the fibrin matrix following the cleavage event. In addition, the re-arrangement of the ternary complex and the release of plasmin are faster in the absence of glycosylation at 184 and in dimerisation. These three factors together increase the turnover rate of tPA.

6. Conclusion

This chapter has confirmed previous findings that tPA expressed in different cell lines has different activities and glycosylation patterns. The glycoform profiles of tPA expressed in different cell lines have been defined at the glycoprotein level without the need to release the oligosaccharides and, in addition, lysine-sepharose affinity chromatography has been used to separate different subsets of glycoforms. Kinetic analysis of these subsets indicated that increasing fibrin stimulated activity correlated with increased affinity for lysine. A fast acting tPA which bound tightly to lysine sepharose was designated type D tPA and partial characterisation suggested that this was a dimer composed of a specific subset of tPA type II glycoforms. Molecular modelling illustrated that such a dimer might form through an inter-molecular disulphide bond at Cys83. The kinetics of the interaction between types I, II and D tPA and types 1 and 2 plasminogen showed a four fold range of activity and the combination of the least glycosylated forms of each molecule had the highest rate of fibrin stimulated activity, and the dimer was the most active form of all. Absence of glycosylation of site 184 in tPA was shown to increase the turnover rate (kcat), which involves increasing the affinity of tPA for fibrin and increasing the rate at which the cleavage event and the release of plasmin occurs. Occupancy of site 289 in plasminogen decreased the rate of the formation of the ternary complex (Km), probably by sterically hindering the binding of plasminogen to tPA. Variable site occupancy of site 184 in tPA has been shown by others to modulate the conversion of single chain to the more active 2-chain form and to modulate the binding of tPA to fibrin, and our results indicate that, in addition, absence of glycosylation at this site may be a pre-requisite for dimerisation. It has been shown that variable site occupancy of site 184 is a means of controlling the turn over rate, confirming the critical role of this glycosylation site in the function of the enzyme. In addition some evidence has been put forward for the existence of a subset of glycoforms which have increased levels of fibrin independent and fibrin dependent activity, and for a role for the neutral and sulphated sugars at site 448 in allowing increased fibrin dependent activity compared with sialylated glycans.

The future

The glycosylation of tPA does not affect the rate of formation of the ternary complex which therefore does not depend on tPA binding to lysine. The observed differences in activity between different sources of tPA and different subsets of glycoforms are related to the turnover rate: One factor which influences this may be the affinity of tPA for fibrin. Binding to fibrin through lysine in K2 is increased in the absence of glycosylation at site 184, and by the presence of oligomannose rather than complex sugars at this site. The glycosylation at other sites does not affect this binding. The effect of glycosylation on fibrin binding through the finger domain, which does not involve lysine, is not known. In addition, another important factor in the turn over rate is the efficiency with which tPA cleaves plasminogen immobilised on fibrin. The role that tPA glycosylation plays in this event has not been well studied. The mechanism of fibrin independent activity is also not well studied; in particular it is not known whether all or some glycoforms of tPA can engage in both fibrin dependent and independent activity or whether there are specific subsets for each. In this respect, analysis of the glycosylation of the minor subset of glycoforms, type Ia, which has low levels of fibrin dependent activity, may give some insight into the role of glycosylation in the fibrin independent activity of tPA. The exact nature of the glycosylation of multidomain proteins may be an important feature of control, particularly for enzymes which are involved in both inter- and intra- molecular rearrangements. In addition to its role in fibrinolysis, tPA can initiate the protease cascade which is associated with remodelling of the extracellular matrix (Stack et al 1990), and in which other glycosylated molecules such as collagenase, stromelysin and gelatinase are involved (Opdenakker and van Damme 1992). Since there is a 2-fold increase in fibrin dependent activity between type II tPA/type 2 plasminogen over type I tPA/type1 plasminogen it is interesting to speculate that further amplification of the range of activity might result from the interaction of specific sets of glycosylated variants of all of the enzymes. If it proves to be so then differential glycosylation would allow the protein a means to modulate the rate of such a cascade. The localised generation of tPA in the ECM is under the control of several ECM-associated proteins which stimulate plasminogen in the absence of fibrin. These include type IV collagen, which is a more efficient stimulator of tPA than fibrin (Stack et al 1990). In this respect, human gelatinase B, which degrades type IV collagen, has been isolated from neutrophils and tumor cells (Professor G. Opdenakker, The Rega Institute). The glycosylation of neutrophil gelatinase B has been defined (ch.8) as a preliminary step towards investigating the possibility that its enzymatic activity may be modulated by glycosylation thus identifying other steps in the cascade process where glycosylation can result in subtle control.

6. References to chapter 7:

Allen, S., Naim, H.Y. and Bullied, N.J. (1995) J. Biol. Chem. 270 4797-4804 Intracellular folding of tissue type plasminogen activator; effects of disulphide bond formation on N-linked glycosylation and secretion.

Astrup, T., and Permin, P.H., (1947) Nature 159, 681-682 Fibrinolysis in the animal organism.

Baron, M., Norman, D., Willis, A. and Campbell, I.D. (1990) Nature 345, 642-646 Structure and function of the fibronectin type 1 module.

Bennett, W.F., Paoni, N.F., Keyt, B.A., Botstein, D., Jones, A.J.S., Presta, L., Wurm, F.M. and Zoller, M.J. (1991) J.Biol.Chem. 266 5191-5201 High resolution analysis of functional determinants on human tissue-type plasminogen activator.

Berg, D.T., Burck, P.J., Berg, D.H. and Ginnell, B.W. (1993) Blood 81, 1312-1322 Kringle glycosylation in a modified human tissue plasminogen activator improves functional properties.

Boose, J.A., Kuismanen, E., Gerard, R., Sambrook, J. and Gething, M.J. (1989) Biochemistry 28 635 The single chain form of tissue type plasminogen activator has catalytic activity: studies with a mutant enzyme which lacks the cleavage site.

Brockway, W. and Castellino, F.J. (1972) Arch. Biochem Biophys 151 194-199 Measurement of the binding of antifibrinolytic amino acids to various plasminogens.

Chan, A.L., Morris, H.R., Panico, M., Etienne, A.T., Rogers, M.E., Gaffney, P., Creighton-Kempsford and Dell, A. (1991) Glycobiology 1 173-185 A novel sialylated N-acetyl galactosamine-containing oligosaccharide is the major complex type structure present in Bowes melanoma tissue plasminogen activator.

Cole, E.S., Nichols, E.H., Poisson, L., Hamois, M.L. and Livingstone D.J. (1993) Fibrinolysis 7 15-22 *In vivo* clearance of tPA: the complex role of sites of glycosylation and level of sialylation.

Collen, D. and Lijnen, H.R. (1991) Blood 78 3114-3124 Basic and clinical aspects of fibrinolysis and thrombolysis.

Cooke, R.M., Wilkinson, A.J., Baron, M., Pastore, A., Campbell, I.D., Gregory, H. and Sheard, B. (1987) Nature 327 339-341 The solution structure of human growth factor.

Dano, K., Andreasen, P., Grondahl-Hansen, J., Kristensen, P., Nielsen, L., and Skriver, L. (1985) Adv. Cancer Res. 44 139-266 Plasminogen activators, tissue degradation and cancer.

Drapier, J.C., Tenu, J.P., Lemaire, G. and Petit, J.F. (1979) Biochimie 61 463-471 Regulation of plasminogen activator secretion in mouse peritoneal macrophages.

Erinarsson, M., Brandt, J., and Kaplan, L. (1985) Biochemica et Biophysica Acta 830 1-10 Large scale purification of human tissue-type plasminogen activator using monoclonal antibodies.

Geppert, A.G. and Binder, B.R. (1992) Arch. Biochem. and Biophys. 297 205-212 Allosteric regulation of tPA-mediated plasminogen activation by a modifier mechanism: evidence for a binding site for plasminogen on the tPA A-chain.

Gonzalez-Gronow, M., Edelberg, J.M., and Pizzo, S.V. (1989) Biochemistry 28 2374-2377 Further characterisation of the cellular plasminogen binding site: evidence that plasminogen 2 and lipoprotein α compete for the same site.

Goochee, C.F., Gramer, M.J., Anderson, D.C., Bahr, J.B., and Rasmussen, J.R. (1992) Frontiers in Bioprocessing II, pp 199-240 Eds Todd, P., Sikdar, S.K. and Bier, M. Pubs. A.C.S. The oligosaccharides of glycoproteins: factors affecting their synthesis and their influence on glycoprotein properties.

Goochee, C.F. and Monica, T. (1990) Bio/Technology 8 421-427

Hajjar, K.A. and Reynolds, C.M. (1993) J. Clin. Invest. 93 703-710 α-fucose-mediated binding and degradation of tissue plasminogen activator by HepG2 cells.

Hall, S.W., Vandenberg, S.R. and Gonias, S.L. (1990) J. Cell Biochem. 43 213-227 Plasminogen carbohydrate side chains in receptor binding and enzyme activation: a study of C6 glioma cells and primary cultures of rat hepatocytes.

Harris, R.J., and Spellman M.W. (1993) Glycobiology 3 219-224 O-linked fucose and other post translational modifications unique to EGF modules.

Harris, R.J., Leonard, C.K., Guzzetta, A.W. and Spellman, M.W. (1991) Biochem. 30 2311-2314 Tissue Plasminogen Activator Has an O-linked Fucose Attached to Threonine-61 in the Epidermal Growth Factor Domain.

Hayes, M.L., and Castellino, F.J. (1979) J. Biol. Chem. 254 8768-8771 Carbohydrate of the human plasminogen variants I.

Hayes, M.L., and Castellino, F.J. (1979) J. Biol. Chem. 254 8772-8776 Carbohydrate of the human plasminogen variants II.

Hayes, M.L., and Castellino, F.J. (1979) J. Biol. Chem. 254 8777-8780 Carbohydrate of the human plasminogen variants III.

Horrevoets, A.J.G., Smilde. A., de Vries, C., and Pannekoek, H. (1994) J. Biol. Chem. 269 12639-12644 The specific roles of finger and Kringle 2 domains of Tissue type plasminogen activator during *in vivo* fibrinolysis.

Hotchkiss, A., Refino, C.J., Leonard, C.K., O'Connor, J.V., Crowley, C., McCabe, J., Tate, K., Nakamura, G., Powers, D., Levinson, A., Mohier, M., and Spellman, M.W. (1988) Thromb. haemostasis 60 255-261 The influence of carbohydrate structure on the clearance of recombinant tissue type plasminogen activator.

Howard S.C., Wittwer, A.J., and Welply, J.K. (1991) Glycobiology 1 411-417 Oligosaccharides at each site make structure dependent contributions to biological properties of human tissue plasminogen activator.

Krause, J., Seydel, W., Heinzel, G., Tanswell, P. (1990) Biochem. J. 267, 647-652 Different receptors mediate the hepatic catabolism of tissue type plasminogen activator and urokinase.

Lee, Y.C., Townsend, R.R. and Hardy, M.R. (1983) J. Biol. Chem. 258 199-202 Binding of synthetic oligosaccharides to the hepatic Gal/GalNAc lectin.

Lijnen, H.R., Van Hoef, B., and Collen, D. (1981) Eur. J. Biochem. 120 149-154 On the role of the carbohydrate side chains of plasminogen in its interaction with α-2 antiplasmin and fibrin.

Madison, E.L., Kobe, A., Gething, M-J., Sambrook, J.F. and Goldsmith, E. (1993) Science 262 419-421 Converting tissue plasminogen activator to a zymogen: a regulatory triad of Asp-His-Ser.

Mori, K., Dwek, R.A., Downing, A.K., Opdenakker, G., and Rudd, P.M. (1995) J. Biol. Chem. 270 The activation of type 1 and type 2 plasminogen by type I and type II tissue plasminogen activator.

Nishimura H, Takao T, Hase S, Shimonishi Y, and Iwananga S. (1992) J. Biol. Chem. 267 17520-17525 Human Factor IX has a tetrasaccharide "O"-glycosidically linked to Serine 61 through the Fucose Residue.

Opdenakker. G., and Van Damme, J. (1994) Immunol. Today 15 103-107 Cytokine regulated proteases in autoimmune diseases.

Opdenakker. G., Bosman, F., Decock, B., Cabeza-Arvelaiz, Van Damme, J., and Billiau, A. (1988) FEBS Lett. 238 1 129-134 Heterogeneity of human tissue-type plasminogen activator.

Opdenakker, G., van Damme, J., Bosman, F., Billiau, A., and de Somer, P. (1986) Proc. Exp. Med. 182 248-257 Influence of carbohydrate side chains on the activity of tissue-type plasminogen activator.

Opdenakker, G., Rudd, P.M. Ponting, C.J. and Dwek, R.A. (1993) FASEB 14 1330-1337 Concepts and Principles of Glycobiology.

Owensby, D.A., Sobel, B.E. Schwartz, A.L. (1988) J. Biol. Chem. 263 10587-10594 Receptor mediated endocytosis of tissue-type plasminogen activator by the human hepatoma cell line HepG2.

Paoni, N.F., Chow, A., Pena, L.C., Keyt, B.A., Zoller, M.J. and Bennett, W.F. (1993) Protein Engineering 6 529-534 Making tissue plasminogen activator more fibrin specific.

Parekh, R.B., Dwek, R.A., Rudd, P.M., Thomas, J.R., Rademacher, T.W. Warren, T., Wun, T-C., Hebert, B., Reitz, B., Palmier, M., Ramabhadran, T. and Teimeir, D.C. (1989b) Biochemistry 28 7670-7679 N-glycosylation and *in vitro* enzymatic activity of human recombinant tissue plasminogen activator expressed in chinese hamster ovary cells and a murine cell line.

R.B. Parekh, R.A. Dwek, J.R. Thomas, G. Opdenakker, and T. Rademacher; A.W. Wittwer, S.H. Howard, R. Nelson, N. Siegel, M.G. Jennings, N. K. Harakas and J. Feder (1989a) Biochemistry 28, 7644-7661 Cell-Type Specific and Site Specific N-Glycosylation of Type I and Type II Human Tissue Plasminogen Activator.

Pirie-Shepherd, S.R., Jett, E.A., Andon, N.L. and Pizzo, S.V. (1995) J. Biol. Chem. 270 5877-5881 Sialic acid content of plasminogen 2 glycoforms as a regulator of fibrinolytic activity.

Pohl, G., Kallstrom, M., Bergsdorf, N., Wallen, P., and Jornvall, H. (1984) Biochemistry 23 3701-3707 Tissue plasminogen activator: peptide analyses confirm an indirectly derived amino

acid sequence, identify the active site serine residue, establish glycosylation sites and localise variant differences.

Ponting, C.P., Marshall, J.M., and Cederholm-Williams, S.A. (1992) Blood Coag. and Fibrin. 3 605-614 Plasminogen: a structural review.

Ranby, M., Bergsdorf, N., Pohl, G. and Wallen, P. (1982) FEBS Lett. 146 2 289-292 Isolation of two variants of native one chain tissue plasminogen activator.

Rijken, D.C., and Collen, D. (1981) J. Biol. Chem. 256 7035-7041 Purification and characterisation of the plasminogen activator secreted by human melanoma cells in culture.

Rijken, D.C., Otter, M., Kuiper, J. van Berkel, T.J.C. (1990) Thrombosis Res. 1990 Supplement X, 63-71

Rijken, D.C., Wijngaards, G., Zaal-de Jong, M. and Welbergen, J. (1979) Biochim. Biophys. Acta 580 140-153 Purification and partial characterisation of plasminogen activator from human uterine tissue.

Roche, P.C., Campbeau, J.D. and Shaw, S.T. (1983) Biochim. Biophys. Acta 745 82-89 Comparative electrophoretic analysis of human and porcine plasminogen activators in SDS polyacrylamide gels containing plasminogen and casein.

Sappino, A.-P., Madani, R., Huarte, J., Belin, D. and Kiss, J.Z., Wohlwend, A. and Vassali, J.-D. (1993) J. Clin. Invest. 92 679-685 Extracellular proteolysis in the adult murine brain.

Smedsrod, B. and Einarsson, M. (1990) Thromb. Haem. 63, 60-66 Clearance of tissue plasminogen activator by mannose and galactose receptors in the liver.

Smedsrod, B., Einarsson, M. and Pertoft, H. (1988) Thrombosis and Haemostasis 59 480-484 Tissue plasminogen activator is endocytosed by mannose and galactose receptors of rat liver cells.

Sottrup-Jensen L., Claeys H, Zajdel M., Petersen ,T.E., Magnussen S. (1978) In Davidson, T.F., Rowan, R.M., Samama, M.M., Desnoyers, P.C. (eds): *Fibrinolysis and Thrombolysis* Vol. 3 New York Raven Press 191-209

Spellman, M.W., Basa, L.J., Leonard, C.K., Chakel, J.A., O'Connor, J.V., Wilson, S., van Halbeek, H. (1989) J. Biol. Chem. 264, 14100-14111 Carbohydrate structures of tissue plasminogen activator expressed in chinese hamster ovary cells.

Stack, M. S., Gonzalez-Grownow, M. and Pizzo, S.V. (1990) Biochemistry 29 4966-4070 Regulation of plasminogen activation by components of the extracellular matrix.

Suenson, E., Lutzen, O., and Thorsen, S. (1984) Eur. J. Biochem. 140, 513-522 Initial plasmin degradation of fibrin as the basis of a positive feed back mechanism in fibrinolysis.

Tanswell, P., Schluter, M., and Krause, J. (1989) Fibrinolysis 3 79-84 Pharmacokinetics and isolated liver perfusion of carbohydrate modified recombinant tissue type plasminogen activator.

Tate, K.M., Higgins, D.L., Holmes, W.E., Winkler, M.E., Heyneker, H.L. and Vehar, G.A. (1987) Biochemistry, 26 338-343 Functional role of proteolytic cleavage at Arginine-275 of human tissue plasminogen activator as assessed by site directed mutagenesis.

Vassali, J.D., Sappino, A.-P. and Belin, D. (1991) J. Clin. Invest. 88 1067-1072 The plasminogen/plasmin system.

Verheijen, J.H., Nieuwenhuizen, W., Traas, D.W., Chang, G.T.G., and Hoegee, E. (1983) Thromb. Res. 32, 87-92 Differences in effects of fibrin(ogen) fragments on the activation of 1-Glu-plasminogen and 442-val-plasminogen by tissue type plasminogen activator.

Verheijen, J.H., Mullaart, E., Chang, G.T.G., Kluft, C. and Wingaards, G. (1982) Thromb. Haemostas. 48 266-269 A simple, sensitive spectrophotometric assay for extrinsic (tissue-type) plasminogen activator applicable to measurements in plasma.

de Vos, A.M., Ultsh, M.H., Kelley, R.F., Padmanabhan, K., Tulinsky, A., Westbrook, M.L. and Kossiakoff, A.A. (1992) Biochemistry 31 270-279 Crystal Structure of the Kringle 2 domain of tissue plasminogen activator at 2.4A resolution.

Wallen, P., Pohl, G., Bergsdorf, N., Ranby, M., Ny, T. and Jornvall, H. (1983) Eur. J. Biochem. 152 681-686 Purification and characterisation of a melanoma cell plasminogen activator.

Walter, J., Steigmann, W., Singh, T.P., Bartunik, H., Bode, W. and Huber, R. (1982) Acta Cryst. B, 38, 1462 On the disordered activation domain in trypsinogen: chemical labelling and low temperature crystallography.

Wittwer, A.J., Howard, S.C., Carr, L.S., Harakas, N.K., and Feder, J., Parekh, R.B., Rudd, P.M., Dwek, R.A., and Rademacher, T.W. (1989) Biochemistry 28 7662-7669 Effects of N-

glycosylation on *in vitro* activity of Bowes Melanoma and human colon fibroblast derived tissue plasminogen activator.

Wittwer, A.J.and Howard, S.C. (1990) Biochemistry 29 4175-4180 Glycosylation at Asn-184 inhibits the conversion of single-chain to two-chain tissue type plasminogen activator by plasmin.

Woods, R.J., Edge, C.J., and Dwek, R.A. (1994) Nature Structural Biology 1 499-501 Protein surface oligosaccharides and protein function.

Zamarron, C., Lijnen, H.R., and Collen, D. (1984) J. Biol. Chem. 259 2080-2083 Kinetics of activation of plasminogen by natural and recombinant tissue-type plasminogen activator.

van Zonneveld, A.J., Veerman, H., and Pannekoek, H. (1986) J. Biol. Chem. 261 14214-14218. On the interaction of the finger and Kringle-2 domain of Tissue-type Plasminogen Activator with fibrin.

Chapter 8

Methods for oligosaccharide and glycoform analysis used in this thesis

Publications associated with this chapter:

1. Rudd, P.M. (1994) Kaseaa (Japan) 32(10) 619-688 Strategies for carbohydrate sequencing.
2. Harvey, D.J. and Rudd, P.M.; R.H. Bateman, R.S. Bordoli, K. Howes, J.B. Hoyes, and R.G.Vickers (1994) Organic Mass Spectrometry 29 753-765 Examination of complex oligosaccharides by matrix -assisted laser desorption mass spectrometry on time-of-flight and magnetic sector instruments.
3. Rudd, P.M., Scragg, I.G., Coghill, E. and Dwek, R.A. (1992) Glycoconjugate Journal 9 86-91 The Separation and Analysis of the Glycoform Populations of Ribonuclease B using Capillary Electrophoresis.
4. Rudd, P.M., Mattu, T. and Honda, S. (1995 *in press*) in 'Capillary Electrophoresis: A Practical Approach' (OUP) Ed. D. Goodall. Capillary Electrophoresis of Oligosaccharides, Glycoproteins and Glycopeptides.

Acknowledgements:

1. The application of MALDI MS to simultaneous oligosaccharide sequencing was developed in conjunction with Dr. David Harvey.
2. HPLC technology was developed by Mr. Geoffrey Guile.
3. The application of Capillary Electrophoresis to the separation of glycoforms was developed in conjunction with Ms. Eva Coghill and Mr. Oleg Iourin.
4. Gelatinase B was prepared by Dr. Stefan Masure (The Rega Institute, Leuven)

Abbreviations:

ANTS: 8-aminonapthalene 1,3,6-trisulphonate; CE: Capillary Electrophoresis; MALDI MS: Matrix Assisted Laser Desorption/Ionisation Mass Spectrometry; RAAM: Reagent Array Analysis Method; FACE: Fluorophore Assisted Carbohydrate Electrophoresis; P4 GPC: P-4 gel permeation chromatography; GPI : glycosylphosphatidyl inositol; WAX: weak anion exchange; APAM: *Aspergillus phoenicus* α1-2 mannosidase; JBAM: Jack bean α–mannosidase; RI: refractive index; gu: glucose units; H: hexose; F: fucose; N: N-acetyl hexosamine; RNase: ribonuclease

Abstract:

In this thesis standard procedures have been used to release and sequence oligosaccharides and these methods are collected together in this chapter. In addition, technology which has been developed recently in the Glycobiology Institute has been used to explore some novel approaches to sequencing. These approaches include the development of a method to sequence mixtures of glycans simultaneously using MALDI MS, a method which eliminates the need to purify individual glycans before using exogiycosidase enzymes. Another application of MALDI MS involved the analysis of the products of the Reagent Array Analysis Method (RAAM) enzyme digests. This method eliminates the need to quantify the products of the enzyme digests, a step which is required in the conventional RAAM procedure. Also in this chapter, recently developed HPLC methods, using both ion exchange and normal phase, have been applied to the analysis of fluorescent and radiolabelled anionic and neutral sugars. There was a significant improvement in the quality of the resolution which could be achieved, both of oligosaccharides carrying different numbers of charged groups and of glycans with the same charge but with different molecular volumes. The composition of the pool of oligosaccharides released from a glycoprotein is commonly used to infer the composition of the original glycoform populations. This indirect method is used partly because relatively few techniques are available for separating and analysing glycoforms directly at the protein level. In this chapter capillary electrophoresis has been used to resolve both the neutral glycoform populations of RNase B and the charged glycoform populations of transferrin in molar proportions. Finally, current technology available for analysing O-linked sugars has been used in a study of the glycosylation of neutrophil gelatinase B. While the N-linked sugars were identified as complex biantennary glycans, the analysis revealed that 85% of the sugars associated with this protein are O-linked and these include a series of linear O-linked glycans with a type 1 core. The largest of these, Galβ1-(4,3,6)HexNAcβ1-(2,3,4,6)Galβ1-(4,3,6)HexNAcβ1-(2,4,3,6)Galβ1-(3)(4,6)GalNAc, is an unusual O-linked sugar; O-glycans of this size more often contain branches. The O-glycans may be clustered in a region of amino acids (452-503) which is rich in proline, serine and threonine and is homologous with a region in α2(V) collagen. The analysis highlighted the need to develop further the technology currently available for analysing O-linked glycans.

Methods for oligosaccharide analysis used in this thesis

In this chapter standard methods for oligosaccharide sequencing used throughout this thesis have been collected together. In addition some new applications of Matrix Assisted Laser Desorption Ionisation Mass Spectrometric (MALDI MS), Capillary Electrophoresis (CE), HPLC and Fluorophore Assisted Carbohydrate Electrophoresis (FACE) to oligosaccharide analysis are described. In addition the current technology available for sequencing O-linked glycans has been used in the analysis of the sugars attached to gelatinase B.

1. Standard methods for oligosaccharide sequencing

For a comprehensive review see Annual Reviews in Biochemistry (Dwek et al 1993).

1.1 Release and re-N-acetylation of glycans

Intact glycoproteins (40μg-2mg) were dialysed into 0.1%TFA and lyophilised. Glycans were released by hydrazine (Fig.1) using the GlycoPrep 1000 (Oxford GlycoSystems Ltd) at a temperature of 95°C for maximum recovery of both N- and O-linked sugars or 65°C for the optimal release of O-linked glycans (Patel et al 1993). Hydrazinolysis releases N- and O-linked oligosaccharides non-selectively in their correct molar proportions and, under optimised conditions, in >85% yield. The GlycoPrep 1000 automatically re-N-acetylates the released sugars to reinstate any N-acetyl groups removed from the original glycans by the hydrazinolysis procedure. The acetohydrazone group is then cleaved yielding the intact glycan in the closed ring form and in its original conformation. The glycans are purified from buffers and excess reagents on ion exchange resins and delivered in aqueous solution (2ml).

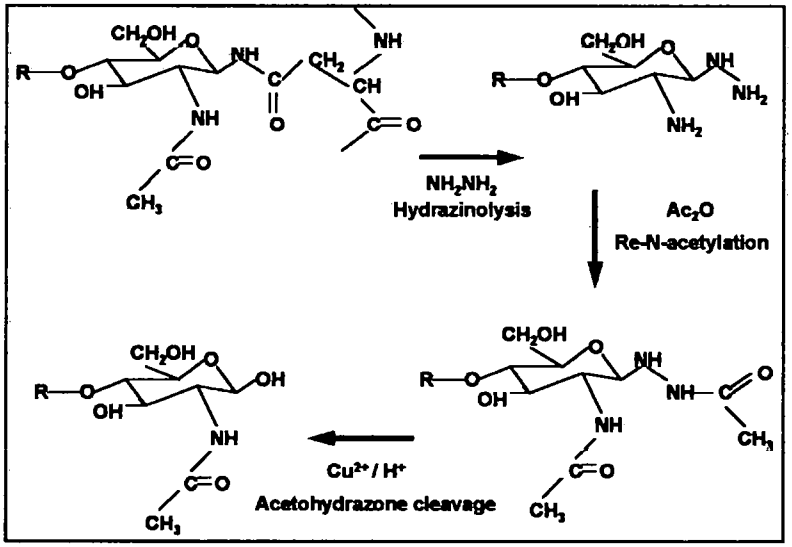


Figure1: Hydrazinolysis reaction followed by re-N-acetylation and acetohydrazone cleavage

1.2 ^3H labelling of reducing terminus

Released and re-N-acetylated oligosaccharides were incubated at 30°C with 2mCi [^{3}H] lactitol and a five-fold excess of NaB^3H_4 (6mM) in 0.05M NaOH adjusted to pH 11 with boric acid. After 4 hours an equal volume of 1M NaB^2H_4 in buffered 0.05M NaOH was added and the incubation continued for 2 hours (Fig.2). The reaction was terminated by the addition of 1M acetic acid and the mixture passed through Dowex AG-50 X12 (H^+) and concentrated to dryness. The residue was freed from boric acid by repeated evaporation (5x) with methanol, dissolved in water and subjected to paper chromatography. After 60 hours, when [^{3}H] lactitol was well resolved from the reduced oligosaccharides the paper was scanned for radioactivity and each radioactive region eluted with water and concentrated (Ashford et al 1987).

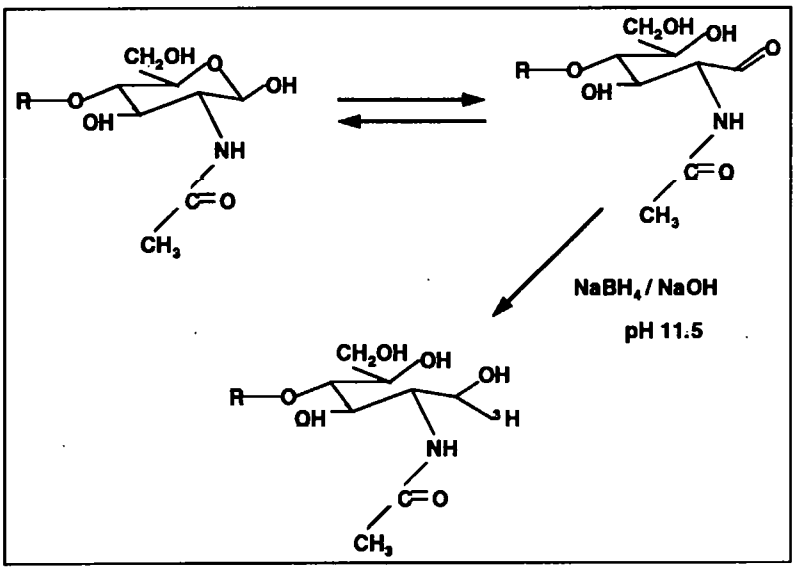


Figure 2: The reactions involved in labelling glycans by reduction with sodium borohydride

1.3 Fluorescent labelling of reducing terminus with 2-aminobenzamide (2AB)

Salt free glycans (25 pmoles - 50 nmoles) were evaporated to dryness using a Savant SpeedVac. 150µl of 2-aminobenzamide were added to DMSO (350µl) and 200 µl of this solution was added to bromophenol blue dye. 100µl of the mixture of 2-aminobenzamide, DMSO and bromophenol blue were added to the reductant (sodium cyanoborohydride) and vortexed. 5µl of the final solution were added to each glycan sample, the tube was capped and the mixture incubated at 65°C for 2 hours (Fig. 3). After the incubation the samples were centrifuged briefly and spotted onto Whatman 3MM paper in a single transfer. A chromatography tank was saturated with butanol:ethanol:water (4:1:1) and the excess reagents were eluted from the labelled glycans by ascending paper chromatography. When the dye reached the top of the paper the strip was removed, thoroughly dried and scanned using a fluorescent instant imager. The 2AB labelled glycans remaining at the origin were eluted with water.

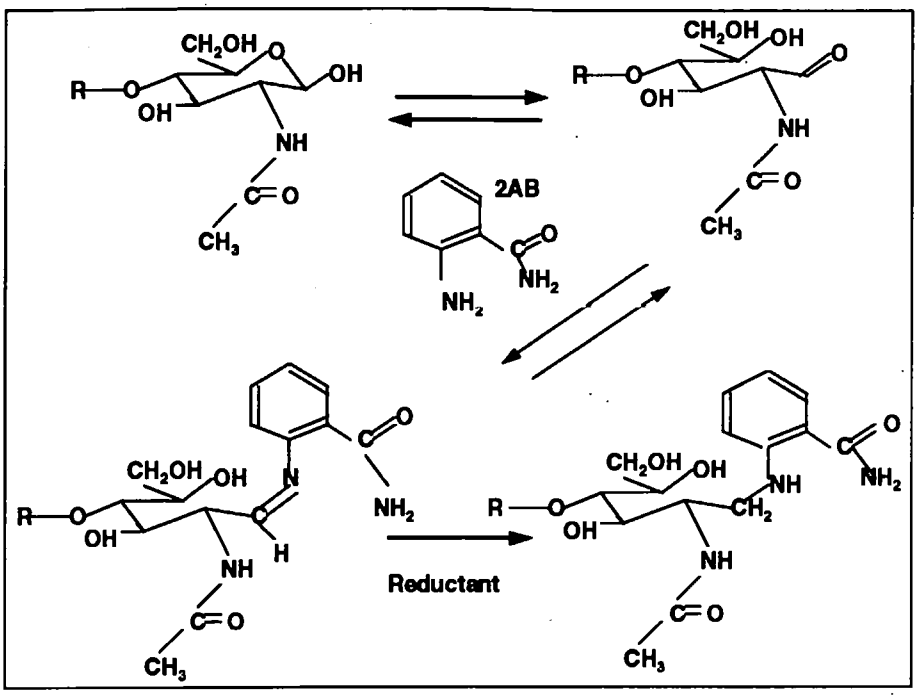


Figure 3: Reductive amination with the fluorescent compound 2AB

(Excitation wavelength: 340nm, emission wavelength 420nm)

1.4 Glycosidase digestions

Glycans were evaporated to dryness on a SpeedVac. 20µl of a standardised solution of the required enzyme were added to the glycans and the mixture incubated for 16h at 37°C. The conditions used for each enzyme are summarised below:

(a) Bacteriodes fragilis endo-β–galactosidase (Oxford GlycoSystems): 0.5U/ml in 50mM sodium actetate buffer pH5.8; substrate concentration 40µM. Cleaves R-GlcNAcβ1-3Galβ1-4GlcNAc/Glc

(b) Arthrobacter ureafaciens neuraminidase (Oxford GlycoSystems): 1-2U/ml in 100mM sodium actetate buffer pH5; substrate concentration 5-30µM. The enzyme cleaves NeuNAcα2-6>3,8 R

(c) Newcastle disease virus neuraminidase (Oxford GlycoSystems): 0.2U/ml in 50mM sodium actetate buffer pH5.5; substrate concentration >5µM. The enzyme cleaves NeuNAcα2-3,8 R

(d) Bovine testes galactosidase (Oxford GlycoSystems): 1-2 U/ml in 100mM citrate/phosphate buffer pH4; substrate concentration 40-100µg/ml. The enzyme cleaves Galβ1-3,4>6 GlcNAc/R.

(e) Streptococcus pneumoniae N-acetyl β-hexoseaminidase: Substrate concentration: 20mM with 300U/ml of enzyme in 100mM sodium citrate phosphate pH6. Incubation: 18h, 37°C. The enzyme cleaves GlcNAc β1-2,4,6 Man or GlcNAcβ1-3,6Gal.

(f) Jack bean N-acetyl β-hexoseaminidase:

Substrate concentration: 20µM sugar/10U/ml enzyme in 100mM sodium citrate buffer pH5 37°C for 16-24 hours. The enzyme cleaves GlcNAc β1-2,4,6R or GalNAcβ1-2,3,4,6R.

(g) Bovine epididymus α-fucosidase (Oxford GlycoSystems) 0.2U/ml in 100mM sodium citrate pH6. Substrate concentration was 40μM. The enzyme cleaves fucose α-1,6>2,>3,>4R.

(h) IgG G0 mix (50μl) contained (i) 15U β-hexoseaminidase (Jack Bean) (300U/ml) (ii) 75mU α-fucosidase (Bovine kidney) (1.5U/ml) in 100mM sodium citrate / phosphate buffer, pH 4.5 containing 0.15M NaCl, 10mM zinc acetate, 0.02% sodium azide, 5mM D-galactonic acid γ-lactone. The last reagent was added to inhibit the activity of galactosidase which may be present as an impurity. 10μl were used for each enzyme digest of approximately 250pmol (0,5mg glycan substrate. The digestion of asialo IgG glycans was carried out overnight at 37°C. After desalting the digestion products were analysed by P4 GPC or by cation exchange chromatography.

1.5 Exoglycosidase sequencing of anchor glycans

In ch.3 the oligosaccharide moiety of the GPI anchor attached to CD59 was analysed using the following enzyme digestion conditions:

(a) Aspergillus phoenicus α1-2mannosidase (APAM) (also known as *A.Saitoi* α1-2 mannosidase.)

This enzyme is specific for α-D-Man1-2D-Man glycosidic bonds. Digestions were performed in 10μl 0.1M sodium acetate pH 5.0 containing 0.01mU (0.04mg) of APAM (Oxford GlycoSystems Ltd. X-5009) for 16-24 hours at 37°C under a toluene atmosphere at a substrate concentration of between 15mM and 2mM. The products were de-salted by passing through 0.2ml AG50X12 (H^+) and eluted with 1ml water. The samples were evaporated to dryness and residual acetic acid removed by flash drying.

(b) Jack bean α–mannosidase (JBAM)

This enzyme has a broad specificity for any terminal, non-substituted, α-D-Man residue.

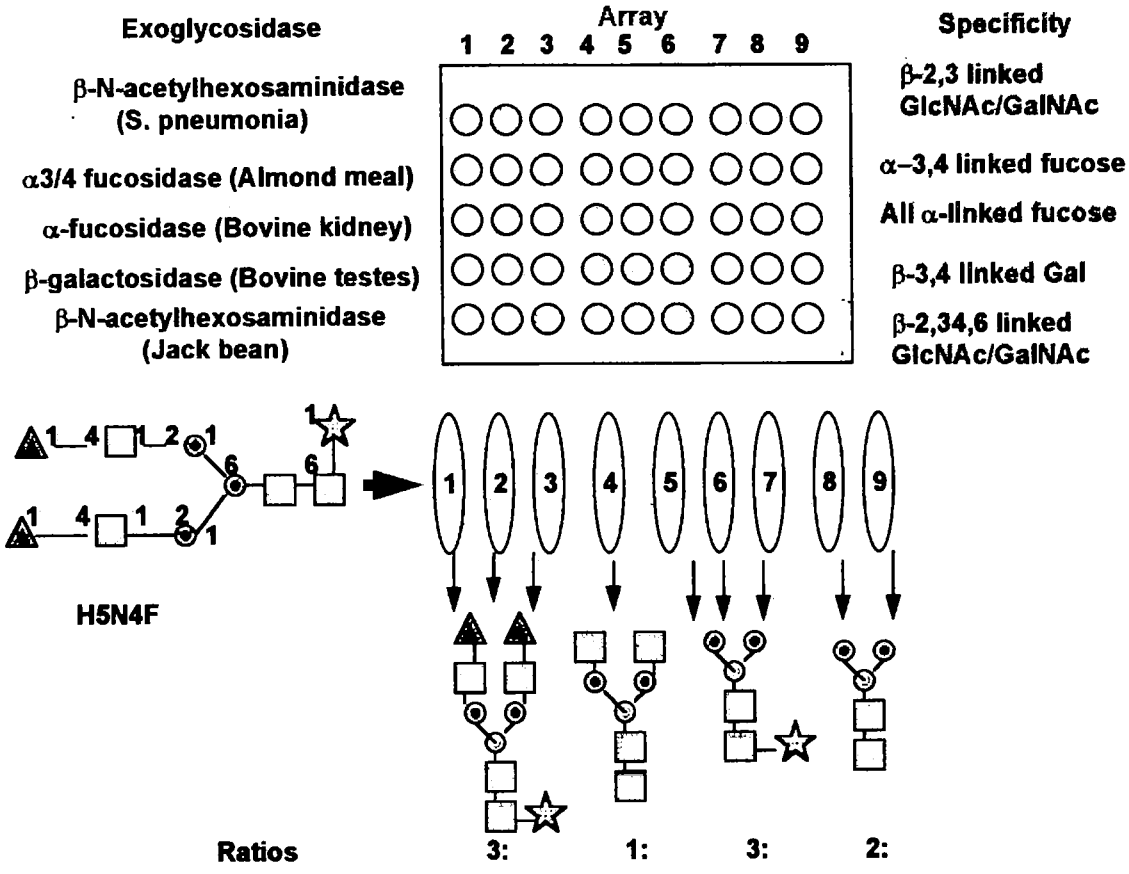
Digestions were performed using 30μl enzyme (25U/ml) in 0.1M sodium acetate pH5. Substrate concentration was between 10mM and 5mM, and the mixture of enzyme and substrate was incubated for 2-4 hours at room temperature followed by 16-24 hours at 37°C. The products were desalted with AG50X12(H^+) (0.2ml), filtered through a 0.2μm membrane and evaporated to dryness. Acetic acid was removed by flash drying with toluene.

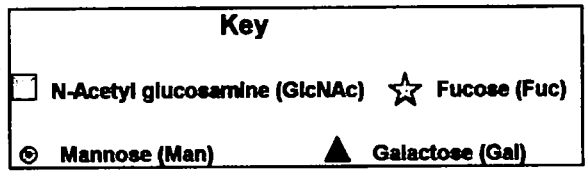
1.6 Purification of enzyme digests prior to P4 analysis and Mass Spectrometry

Samples were purified from protein and salts by passing through mixed bed resins of Chelex100 (Na^+)/Dowex AG50X12 (H^+)/Ag3X4A (OH^-)/QAE Sephadex A-25, eluting with 5-10 bed volumes of water and evaporating to dryness.

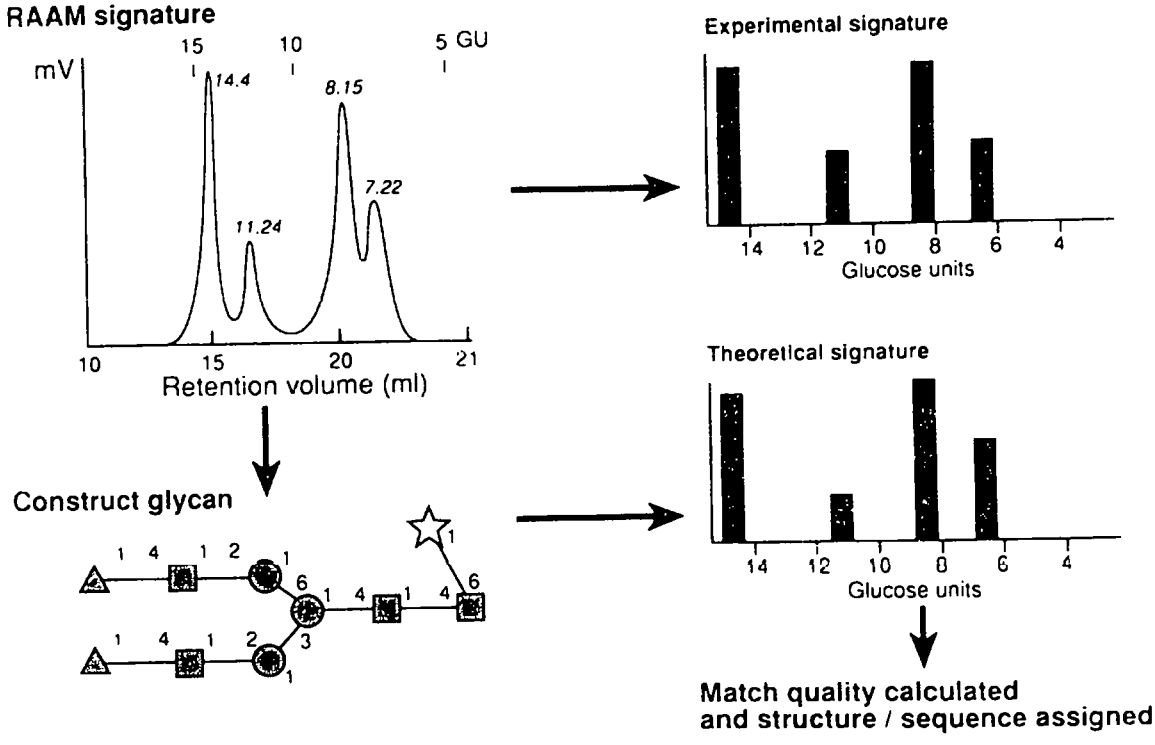
1.7 Biogel P-4 gel filtration

This technique allows neutral glycans to be separated on the basis of hydrodynamic volume, and gives a direct measurement of size. It can be used preparatively, without internal standards, or analytically by co-chromatographing a set of glucose oligomer internal





The Reagent Array Analysis Method



standards with radiolabelled neutral glycans. The simultaneous detection of the oligomer standards by refractive index (RI) and the neutral glycans by an on-line radioactive or fluorescence monitor allows accurate determination of the size of the glycans in terms of effective 'glucose units' (gu). A combination of specific exoglycosidase digestions and re-chromatography of the digests enables the glycan structures to be analysed. The analyses described in this thesis were performed using the Oxford GlycoSystems Glycomap Instrument which is based on the P4 chromatography system.

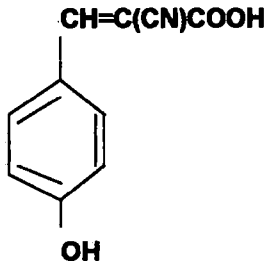
1.8 Reagent Array Analysis Method (RAAM) for Automated Oligosaccharide Sequencing

The RAAM method for the primary sequence analysis of N-linked oligosaccharides (at the picomole level) uses exoglycosidases in multiple defined mixtures (Edge et al 1992, Prime and Dearnley 1993). The analysis is performed in a single chromatographic step. The method, schematically summarised in figure 4a, involves dividing an aqueous oligosaccharide solution, in this case of the complex biantennary glycan H5N4F, into equal aliquots. Each aliquot is incubated with a precisely defined mixture of exoglycosidases, the products of each incubation are combined, and a single analysis is performed on the pool of products. The set of mixtures of exoglycosidases is referred to as the 'reagent array'. Each reagent array is used to generate a fragment of the oligosaccharide, and by omitting one or more different exoglycosidase(s) from each mixture, different 'stop-point' fragments of the oligosaccharide are generated. In each enzyme mix, digestion will continue until a linkage is reached that is resistant to all the exoglycosidases present in that mix. Full use is therefore made of both positive data (the exoglycosidases hydrolyse linkages up to the 'stop' point) and negative data (the exoglycosidases do not hydrolyse linkages beyond the 'stop' point). The original oligosaccharide is labelled at the single reducing terminus, therefore only products retaining the original reducing terminus are seen by the detection system. P4-chromatographic separation of the combined products generates a pattern of fragments that is a 'signature' of that oligosaccharide generated by the specific array used (Figure 4b). This signature is characterised by the elution position and relative signal intensity of each fragment. The software algorithm performs a re-construction of carbohydrate structure from the observed signature (Fig. 4c).

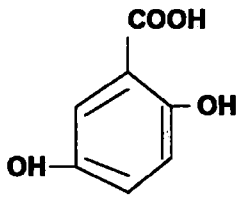
2. Novel applications of technology to oligosaccharide and glycoform analysis

2.1 Matrix Assisted Laser Desorption Mass Spectrometric (MALDI MS)

(i) MALDI MS is a relatively simple mass spectrometric technique (Dwek et al 1993, Harvey et al 1994, Beavis et al 1990, Karas and Hillenkamp 1988, Hillenkamp and Karas 1990) which gives molecular weights without the need to derivatise the oligosaccharides and which does not produce extensive fragmentation. The sample is presented to the ion source co-crystallised with a UV absorbing matrix. The composition of the matrix affects the efficiency of the desorption process; this may be because the energy which the matrix absorbs from the ion source is transmitted to the sample. A number of compounds have been tested as matrices in UV desorpton mass spectrometry of oligosaccharides, in particular cinnamic acid derivatives (Beavis and Chaitt 1990) such as α-cyano-4-hydroxycinnamic acid (Beavis et al 1992) are frequently used. In this thesis the matrix was dihydroxybenzoic acid (DHB) (Stahl et al 1991). The structures of these two compounds are shown below.



α-cyanohydroxycinnamic acid



dihydroxybenzoic acid

Neutral glycans and DHB were first co-crystallised from acetonitrile : water (70:30) and then re-crystallised from ethanol. The dried target was placed in the MS and irradiated with pulses of laser light. The magnetic sector instrument used in this thesis was a Fisons Instruments (VG Analytical) AutoSpec-FPDQ spectrometer consisting of a standard Autospec instrument fitted with an array detector and a LSI nitrogen laser (frequency of 337 nm; pulse width 3nsec). The highest mass which can be measured without significant loss of resolution is 5kD and the accuracy is approximately +/- 0.3 of a mass unit. While proteins and glycopeptides generally give abundant $[M+H]^+$ ions together with low amounts of dimer and doubly charged peaks, oligosaccharides give only single peaks which arise from the $[M+H]^+$ or, more commonly $[M+Na]^+$. The technique requires only fmol-pmol amounts of material and is tolerant of small amounts of buffer salts, although in this thesis all samples were purified by ion exchange chromatography to prevent unnecessary suppression of the signal. The detection limit of the signal was taken to be approximately 3 times the noise level.

Oligosaccharide analyses were performed either on unlabelled neutral glycans or on 2AB labelled glycans. Released glycans were desialylated with *Arthrobacter Ureafaciens* neuraminidase which cleaves both α–2,3- and α–2,6- linked sialic acid. Analysis of radiolabelled glycans before and after neuraminidase treatment confirmed that all charge was due to sialic acid and had been removed. The neutral glycans were purified from

neuraminidase and from buffer salts by passing through a mixed bed resin consisting of 200µl each of Chelex 100 (Na^+) over Dowex AG50X12(H^+) over AG3X4 (OH^-) over QAE Sephadex A-25 and eluting with 3ml water. The sugars were evaporated to dryness by rotary evaporation and dissolved in 10µl water. Approximately 50-100 pmol of sugars in 1-5µl water were twice recrystallised on the target with 0.5µl dihydroxybenzoic acid. 5-20 shots were taken at each exposure and 20-30 spectra were averaged. The molecular weights of the components of the mixtures were determined after calibrating each spectrum with standards. The molecular weight of an oligosaccharide is the sum of its monosaccharide components plus water. The glycans are normally present as sodium adducts and this accounts for a further 23D. Two molecular weights were used (Table 1) depending on the resolution used to obtain the spectra. In high resolution spectra, in which more shots are fired, more sample is required and individual isotopes are identified, the monoisotopic masses of the most abundant isotopes were used; for low resolution spectra in which isotopes are not resolved, the chemical mass was used.

Monosaccharide residue	**Monoisotopic mass**	**Chemical mass**
Hexose	162.053	162.14
Deoxyhexose	146.058	146.14
N-acetyl hexosamine	203.079	203.19
Pentose	132.042	132.12
Sialic acid	291.095	291.26
Water	18.011	18.02
Sodium	22.990	23.00
Potassium	39.10	38.964

Table 1: Monoisotopic and chemical masses of the common components of oligosaccharide ions

Applications

(i) MALDI MS analysis of glycans by simultaneous digestions of oligosaccharides in glycan pools. In this thesis a method was developed to overcome the need to purify individual oligosaccharides before sequencing and to eliminate the need for labelling (Harvey et al 1994). Released glycans were desialylated and desalted as described above and the components analysed by MALDI MS. The glycan pool was then incubated with a single exoglycosidase enzyme and after purification the products of the digest were analysed by MALDI MS. An examination of the profiles showed which sugars had been digested by the enzyme and how many residues had been removed. The sugars were subjected to a series of enzyme digestions in turn and the analysis of the products of each allowed the composition of the original pool to be determined. In chapter 2 this technology was used to

sequence the N-linked glycans released from the leucocyte antigens CD2, CD5d1, CD5/CD4d3/4 and sCD59 and on IgG sugars in chapter 6 (section 4.2.5). Smaller O-linked glycans could not be unambiguously identified by this matrix assisted laser desorption method using DHB because the components of the matrix, which also ionise, have molecular weights in the same range as O-linked glycans (approximately 350-800 daltons). This method therefore requires modification before it can be applied to O-linked glycans.

A major advantage of the magnetic sector instrument is that the individual peak heights are directly proportional to the ion concentration of each component. In contrast, for spectrometers using time of flight separations, this relationship becomes non-linear at lower molecular weights (<1000 daltons) (Mr. T. Naven, unpublished data). It has also been demonstrated that oligosaccharides are non-selectively ionised (Dr. D. Harvey - unpublished data). However, to obtain quantitative data it is necessary to take many scans (typically 20-30) over a wide area of the sample spot, and there may still be some doubt in the quantitation since this depends on an even distribution of the individual sugars across the target area. In the following section a method was developed to analyse RAAM data by mass spectrometry which did not rely on accurate quantitation.

(ii) MALDI MS analysis of RAAM sequencing data of complex N-linked sugars

A method was developed to modify the RAAM sequencing technique (section 1.8) for use with unlabelled glycans using MALDI MS to analyse the data instead of P4 GPC. The standard RAAM analysis using the N-linked glycan array (shown in Fig. 4a) was applied to 2 separate aliquots of unlabelled desialylated glycans from CD5. At the end of the incubation time one set of samples was pooled and purified on mixed bed resins as in the normal protocol. The second set was not pooled, but each reaction product was individually purified on mixed bed resins. The components of both the pooled sample and the nine individual purified reaction products were analysed by MALDI MS. The stop point fragments are shown schematically in figure 5.

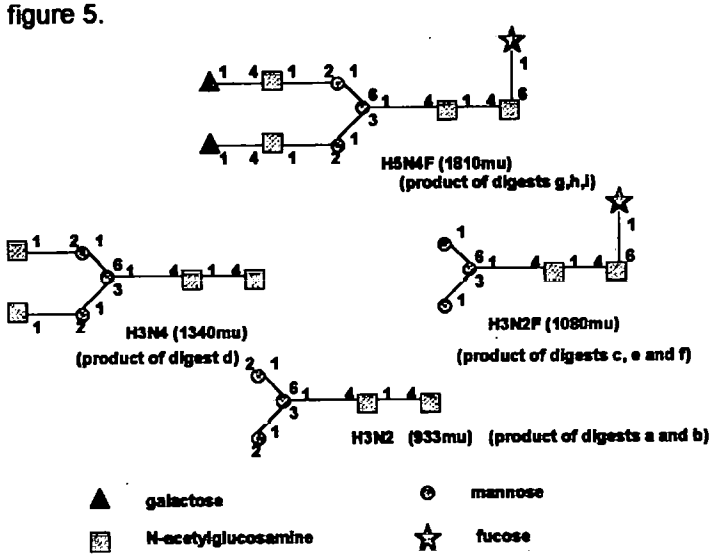


Figure 5: The stop point fragments generated by the nine individual RAAM digests

Predicted ratios of H5N4F:H3N4:H3N2F:H3N2 are 3:1:3:2

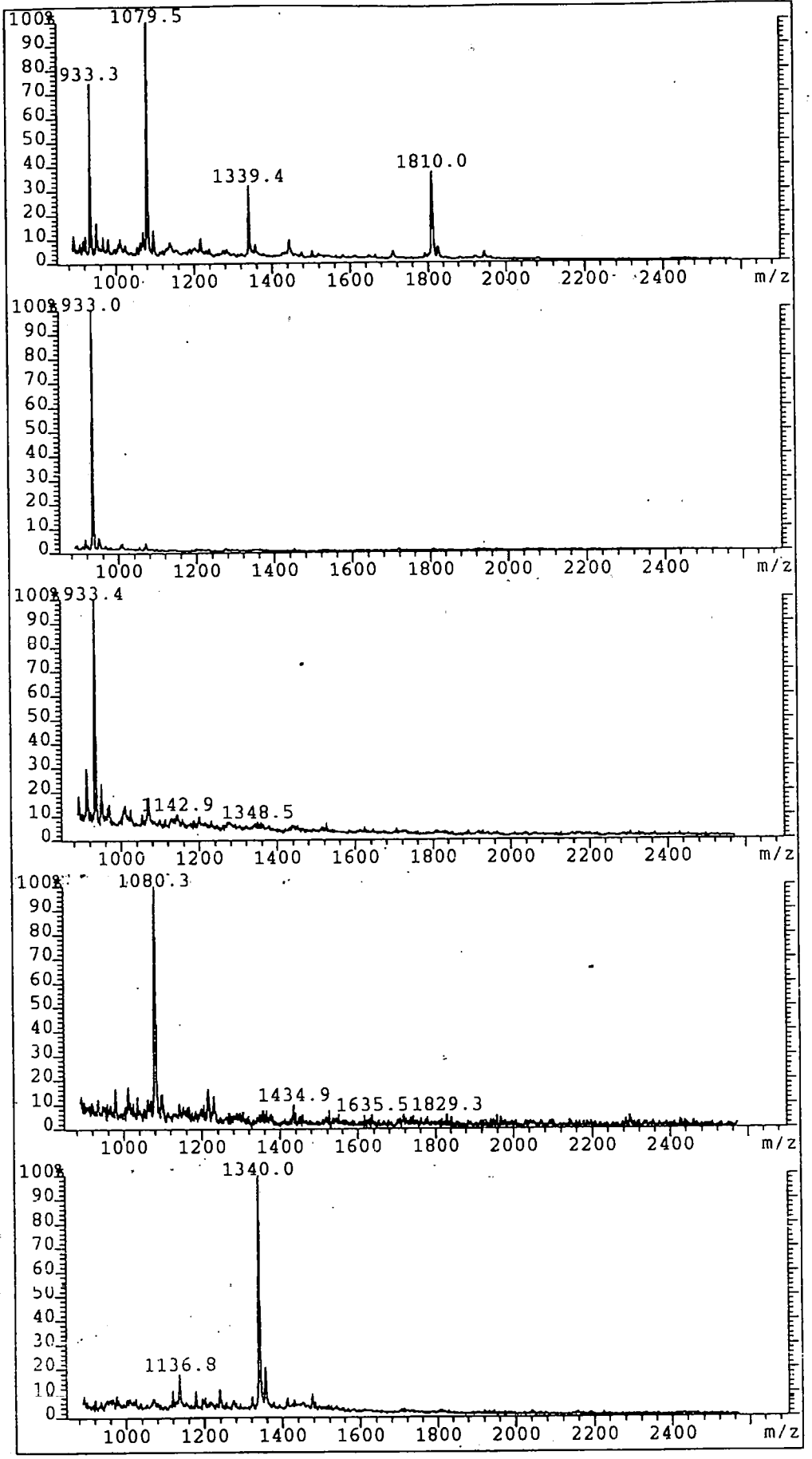


Figure 6P: The MALDI MS analysis of the combined stop point fragments generated from the individual RAAM digests of H5N4F. 6 a-d: The MALDI MS analysis of the individual stop point fragments generated from the individual RAAM digests of H5N4F (continued overleaf). *Data analysis is in tables 2a,b.*

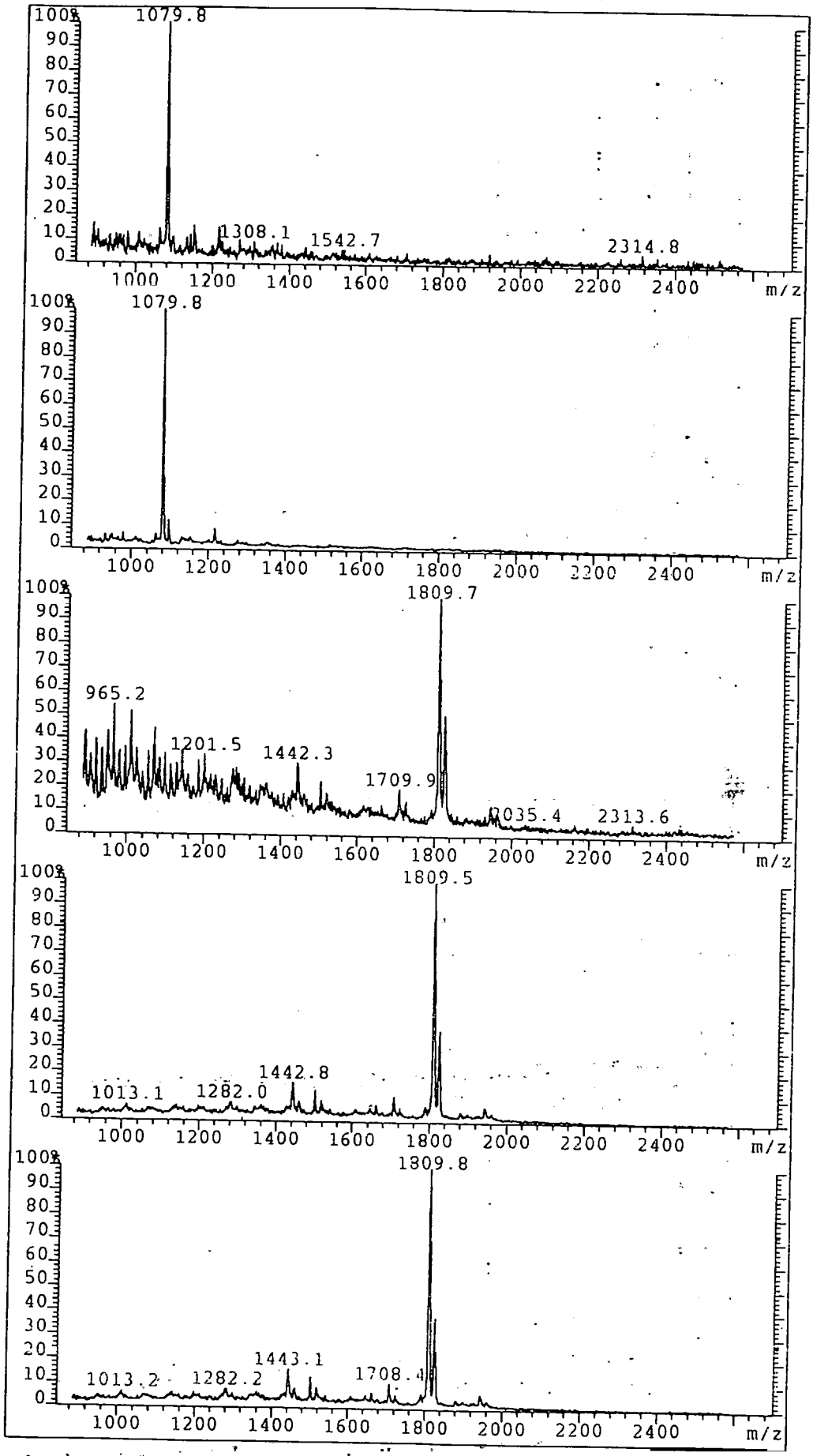


Figure 6 e-i: The MALDI MS analysis of the stop point fragments generated from the individual RAAM digests of H5N4F. *(Data analysis is in Table 2a,b.)*

In the standard RAAM protocol the software algorithm performs a re-construction of the original carbohydrate structure from the observed P4 GPC signature, and this is derived from the quantitation of the relative amounts of stop point fragments in the final pool of glycans. The results of the conventional RAAM analysis of this sample, H5N4F, are shown in figure 4b,c, and the analogous MS analysis of the pooled fractions is shown in figure 6P. It can be seen from Table 2a that the quantitative MS analysis of the RAAM stop point mixture does not correlate well with the predicted ratios (Figure 4a). The peak height for the 1810 mass units fragment (H5N4F) is low, maybe as a result of losses or because of variations in the distribution of this species on the target. To overcome the need to depend on peak height measurements the second set of digests were not pooled but analysed individually. In this case each MS profile (figure 6 a-i) is of a single stop point fragment. The ratios of the products obtained from the MS analysis of these individual incubations are given in Table 2b, together with the theoretical ratios of the stop point fragments. The ratios of the stop point fragments, from which the original carbohydrate can be reconstructed, is obtained by adding up the number of each of the different stop point fragments from the 9 digests (6a-i).

Fig. 6P: MS of pooled digests	933	1081	1340	1810
Ratio of peaks from MS peak height	1.9	2.6	0.8	1
Theoretical ratio of stop point fragments from H5N4F:	2	3	1	3

Table 2a: MS analysis of pooled stop point fragments after RAAM digests compared with theoretical ratio.

6a. RAAM 1	933			
6b. RAAM 2	933			
6c. RAAM 3		1081		
6d. RAAM 4			1341	
6e. RAAM 5		1081		
6f. RAAM 6		1081		
6g. RAAM 7				1810
6h. RAAM 8				1811
6i. RAAM 9				1812
Ratio of peaks by addition of results of individual digests:	2	3	1	3

Table 2b: The ratios of the stop point products derived from the MS analysis of the individual incubations

Table 2: RAAM analysis of the asialo N-linked CD5d1 glycan analysed by MALDI MS

2.2 Ion exchange separation of anionic sugars

In anion-exchange chromatography charged glycans (almost invariably acidic) are eluted sequentially from a suitable column by an increasing concentration of a counter anion. This separation is based on net charge, although in some cases additional interactions between the uncharged body of the glycans and the resin are involved. Two ion exchange systems commonly used for resolving sugars are high performance anion exchange chromatography (HPAEC) using PAD or radioactive detection (Dionex) and mono-Q sepharose anion exchange chromatography. In this thesis weak anion exchange (WAX) chromatography, developed by Mr. Geoffrey Guile (Guile et al 1994) was used to provide a high resolution, preparative method to separate charged glycans without degradation. The technology was applied to the charge analysis of the glycans released from the leucocyte antigens (ch.2) and to the glycans from normal and RA IgG. The improved resolution, which resulted from the use of the fluorescent label, 2AB, and on-line fluorescence detection, revealed that the RA IgG sugars contained two populations of monosialylated glycans compared with one in normal IgG (ch.6 section 4.2.3 figures 10a,b and 10cd). The WAX column (Vydac 301VHP575) (7.5cmx50mm) was equilibrated in water (buffer B) and 2AB or radiolabelled glycans were injected in aqueous solution. Neutral structures were eluted in the void volume and charged glycans which bound to the column were successively eluted with an increasing salt gradient of 0-500mM ammonium formate pH9 (bufferA). The classes of charged glycans were assigned by comparison with a standard fetuin glycan library.

The gradient applied to the column is shown in table 3.

Time	Flow	Curve	%A	%B
Start	1ml/min		0	100
12min	1ml/min	6	5	95
25min	1ml/min	6	21	79
50min	1ml/min	6	80	20
55min	1ml/min	6	100	0
65min	1ml/min	6	100	0
66min	1ml/min	6	0	100
70min	2ml/min	6	0	100
89min	2ml/min	6	0	100
90min	1ml/min	6	0	100
117min	0ml/min	6	0	100

Table 3: Gradient parameters for WAX chromatography (Vydac 301 VHP575)

Typical profiles are shown in **figure 7a,b.**

2.3 Normal phase separation of neutral sugars

This technology was developed to provide a high resolution separation system as a complement to the existing P4 technology. Fluorescently labelled glycans were applied to two tandem Vydac 400 VHP cation exchange columns (2x7.5cmx50mm) equilibrated in water. The oligosaccharides were eluted with a linear gradient in which the starting solution was water containing 50mM formic acid and the final solution was 80% acetonitrile also containing 50mM formic acid. The technology was used to determine the relative proportions of G2:G1:G0 sugars associated with normal and RA IgG (ch.6 section 4.2.6 fig. 15a,b).

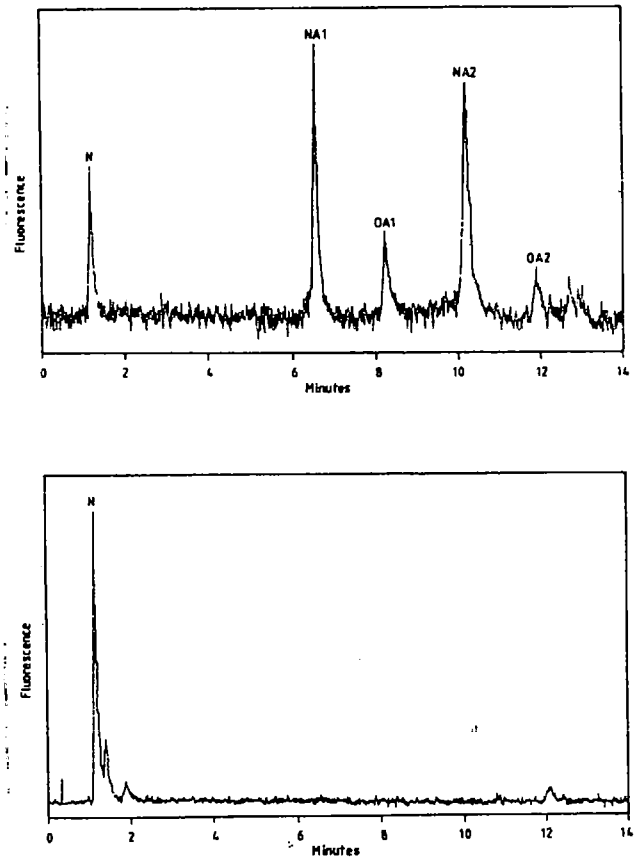


Figure 7: (a) Analysis of the hydrazine released pool of 2AB labelled glycans from CD5d1 by WAX chromatography.

Assignment was by comparison with standard fetuin sugars

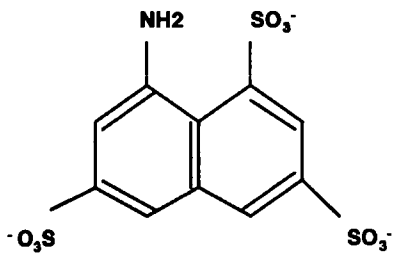
(b) Analysis of 2AB labelled glycans after digestion with Newcastle Disease virus neuraminidase.

All structures became neutral, indicating that all charge is due to 2,3-linked sialic acid.

2.4 Fluorophore Assisted Carbohydrate Electrophoresis (FACE) using Glyco/Millipore electrophoresis and imaging system and labelling kits.

This technique was explored as a means of using a gel electrophoresis system to profile the glycosylation of picomole quantities of glycans. It is based on the principle that sugars

labelled at the reducing terminus by reductive amination with the charged fluorescent group, 8-aminonapthalene-1,3,6-trisulphonic acid (ANTS), will be resolved on 60% acrylamide gels approximately according to their mass : charge ratios (Jackson 1994). The three sulphonic acid groups donate a net negative charge to the labelled glycans allowing both acidic and neutral sugars to be resolved. The fluorescent label enables the derivatives to be detected with high sensitivity (in the picomole range). A wheat starch digest consisting of a mixture of malto-oligosaccharides from 1 to approximately 30 glucose units is provided as a standard marker against which sample mobility can be measured. This technique has proved to be a useful means of making a rapid qualitative assessment of the charge states of oligosaccharides, and of the complexity of the components of glycan libraries released from glycoproteins. Absolute assignment of structure may be attempted through repeated analyses following the use of exoglycosidase enzymes.



ANTS

8-aminonaphthalene-1,3,6-trisulphonate

Method

Hydrazine released oligosaccharides (500pmoles) in water were dried to a pellet in a centrifugal vacuum evaporator and re-dissolved in 5µl of ANTS/acetic acid/water (0.15M/3/17v/v). 5µl of sodium cyanoborohydride dissolved in dimethyl sulphoxide (1M) were added to the mixture which was then vortexed and spun in a microcentrifuge. The samples were incubated at 37°C overnight, or at 45°C for 3 hours, and then evaporated to a viscous gel in a centrifugal vacuum evaporator. The fluorophore labelled sample was resuspended by adding 10-20µl of distilled water. 2µl of the sample (50pmoles) were mixed with 2µl of glycerol/water (1:4 v/v) thorin1 dye and loaded onto the gel. The standard wheat starch digest glycan mixture (1.25pmoles) was suspended in 50µl of distilled water. 50 µl of glycerol /water (1/4 v/v) with thorin1 dye were added and 4µl (50pmole) were applied to the gel. The samples were analysed by SDS PAGE using pre-poured 60% acrylamide gels and 0.1M Tris/Tricine running buffer, pH 8.3. The gels were run at 10W/gel for between 1.5-2 hours and scanned with a fluorescent instant imager.

Application

FACE was used to assess the distribution of charge on the bi-antennary complex glycan, H5N4F, attached to CD5 domain 1 (CD5d1) and to compare the results with the charge analysis by WAX chromatography (Fig. 7a,b). The results of the FACE analysis of the 2AB labelled CD5 glycans are in figure 8a,b. The three bands visible in Track 2 are consistent

Line 2 on page 255:

2AB should read ANTS

with the presence of three populations of N-linked glycans (N, NA1 and NA2), tracks 1 and 3 contain 2AB labelled glucose oligomer standards. In figure 8b (Track 2) the three populations have been reduced to a single band by *Arthrobacter ureafaciens* neuraminidase, consistent with chromatography data (ch.2B section 3.2) which indicates that the charge associated with the CD5d1 glycans is due to sialic acid. Band intensity was determined using instant imaging fluorescent densitometry; the results are summarised in table 4.

The analytical principle of a mass:charge separation is the same for both the WAX and the FACE systems and neither technique can unambiguously assign structures without second dimension analysis. Peak assignment of charged glycans on the WAX HPLC system is normally made by comparing the profile of the sample with that of a standard set of fetuin N- and O-glycans. While a similar approach can be applied to the electrophoresis system, the resolution required for accurate assignment of the elution position of the differently charged species is poor compared with WAX HPLC (figure 7); in addition the interpretation of the data is complicated by the presence of a number of artefactual bands (not assigned) which commonly occur on the FACE gels. Moreover the O-linked sugars are not visible on the FACE gels because the technique, although very sensitive (low p mole range, compared with WAX high p mole range), does not have the range of sensitivity needed to detect the smaller populations of charged O-links (12% and 15% figure 8) at the same time as the more abundant N-linked populations (32% and 36% figure 8). The possibility also exists that highly charged low mass species will elute from the end of the gel during the electrophoresis. To address these problems it would be necessary to compile the data from a range of different sample loadings, run times and gel exposures. Quantitative analysis of individual glycan pools by FACE is generally in line with the data obtained by WAX chromatography (Table 4), but standardisation is needed to make an accurate assessment of the potential for this technique.

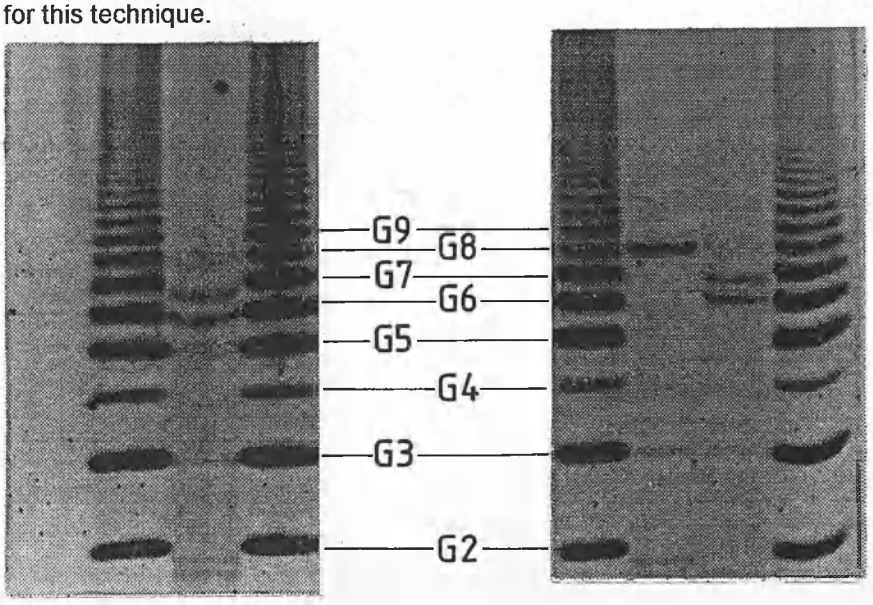


Figure 8: The separation of ANTS labelled CD5d1 glycans by gel electrophoresis (FACE system). (a) *Tracks 1 and 3: ANTS labelled standard mixture of glucose oligomers. Track 2 shows the analysis of hydrazine released glycans from CD5d1, labelled with ANTS.*

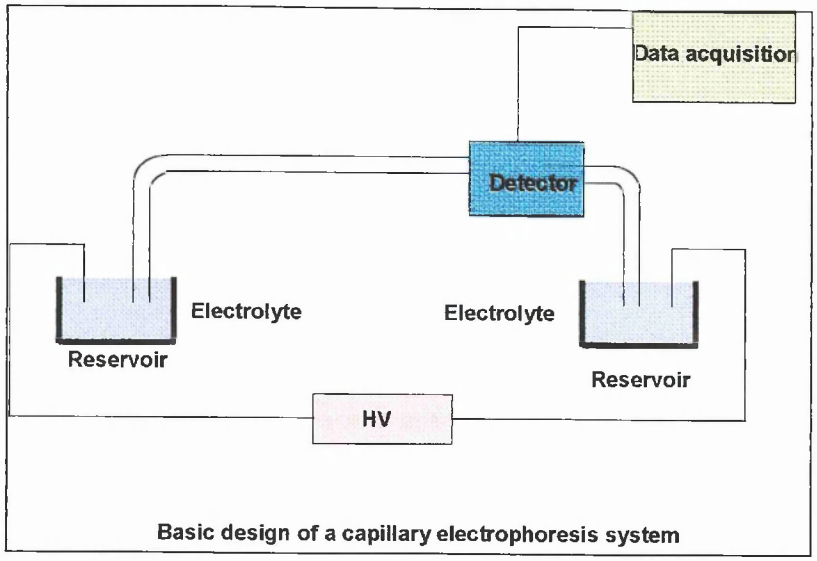


Figure 9a: The basic CE instrument consists of a fused silica capillary with an optical window, a UV detector, a high voltage power supply, two electrode assemblies and two buffer reservoirs. Field strengths higher than 500 volts/cm can be applied to the narrow capillary allowing highly enhanced resolution compared with conventional gel electrophoresis.

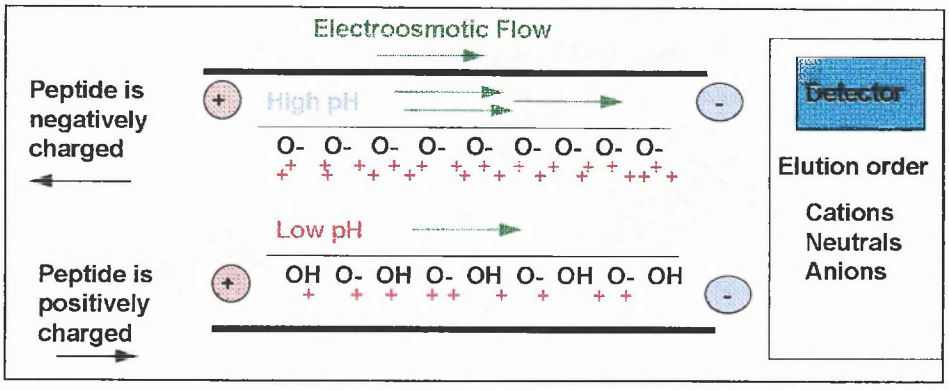


Figure 9b: Effect of electroosmotic flow (EOF). In uncoated capillaries EOF is always towards the cathode. At high pH the capillary walls are strongly negatively charged and the EOF is greater than at low pH. At high pH most peptides are negatively charged and migrate towards the anode, however EOF is always stronger and anions are eventually swept to the cathode. At low pH most peptides are positively charged and move towards the cathode with the EOF. The elution order is the same in both cases: cations elute first, then neutrals and finally anions.

(b) *Analysis of the same glycan pool before (track 3) and after (track 2) desialylation with neuraminidase. Tracks 1 and 4: ANTS labelled standard mixture of glucose oligomers.*

In 8a the bands which have electrophoresed to positions equivalent to 8,7 and 6 glucose oligomers are the neutral, mono- and di-sialylated CD5d1 glycans respectively. Only the major bands were assigned and the O-linked glycans, identified on the WAX chromatography analysis (Fig.7), which have a relatively low abundance and high mass:charge ratio were not identified. The gel has been exposed to the scanner for longer than the gel in 8b to decrease the errors in the densitometry readings for the neutral band. Figure 8b shows the analysis of the same glycan pool before (track 3) and after (track 2) desialylation with neuraminidase. In track 2 the monosialyated glycan has been reduced to the size and charge of the neutral glycan by the loss of one sialic acid residue, while the di-sialylated glycan has been reduced to the size and charge of the neutral glycan by the loss of two sialic acid residues. All three structures visible in figure 8a are therefore located at the same position as the neutral glycan in figure 8b. The bands were assigned from the analysis of the glycans by P4 GPC and WAX chromatography and from the standard elution position of the core fucosylated biantennary complex sugar (H5N4F) on the FACE system.

	N-N-links	N-O-links	NA1	NA2	OA1	OA2
(i) **WAX (7a)**	——15%——		32%	36%	12%	5%
(ii) **WAX (7a/b)**	17%		38%	45%		
(iii) **FACE (8a)**	10%		40%	50%		
	Total N-links	**Total O-links**				
(iv) **WAX (7b)**	85%	15%				
(v) **P4**	66%	33%				

Table 4: A comparison of the charge profiles of CD5 obtained by different methods. (i) The percentages of the charged (NA1, NA2, OA1, OA2) and neutral populations calculated from figure 7a. The percentage of neutral N- and O- links cannot be separated. (ii) The percentages of the N-linked population only, calculated from the charged and asialo WAX profiles (fig. 7a,b) (iii) The N-linked populations analysed by FACE (Fig. 8a) (iv) The ratio of total N : O-links analysed by WAX from figure 7b. Percentages were obtained by integration of the WAX data by fluorescent densitometry of the FACE data. (v) The ratio of N:O-links from the P4 data in ch.2B section 3.3 figure 3.

2.5 The resolution of glycoforms by Capillary Electrophoresis

Capillary electrophoresis (CE) (Jorgensen and Lukacs 1983), which offers the possibility of separating intact glycoforms at the protein level, may be a powerful addition to the existing carbohydrate technologies (Rudd et al 1995). At present other electrophoretic techniques such as SDS PAGE and isoelectric focusing (IEF) generate a limited amount of analytical information relating to intact glycoforms. However the resolution which can be achieved is

restricted because the high voltages which are required to give Improved separations give rise to problems of heat dissipation in conventional gel systems. In CE field strengths higher than 500 volts/cm (up to 30kV) can be applied to a narrow capillary giving the potential for highly enhanced resolution. The basic CE instrument consists of a fused silica capillary with an optical window, a UV detector, a high voltage power supply, two electrode assemblies and two buffer reservoirs (Fig.9a). In most assay systems the inner wall of the capillary column carries a negative charge because the silanol groups ionise above pH3. Positive ions in the electrolyte are attracted to the wall and, under high voltage, migrate towards the cathode creating a bulk flow of liquid called the electroosmotic flow (EOF) (Fig. 9b). If a sample is introduced into the capillary at the anode side all the components are carried by the EOF towards the cathode. Simultaneously the current flow will retard the negatively charged species so that the positive species elute first followed by the neutral and finally the negative. An inherent difficulty associated with analysing glycoforms by CE resides in the fact that the homogeneity of the polypeptide chain frequently masks the diversity of the glycans associated with it. Separations are achieved by amplifying structural differences between the oligosaccharides, for example, by complexing with an ionic species. The chemistry which has been used to achieve separations has, for the most part, exploited subtle differences in the conformation of the hydroxyl groups associated with the monosaccharide residues which make up the oligosaccharides. For example, the oligomannose glycoforms of bovine pancreatic RNase B have been separated into their individual glycoforms in their correct molar proportions by CE (Fig. 10a) (Rudd et al 1992 and ch4) by complexing cis diols with borate ions. The glycoforms of transferrin, also resolved by CE (Fig. 10b) are the subject of ch.5 in this thesis.

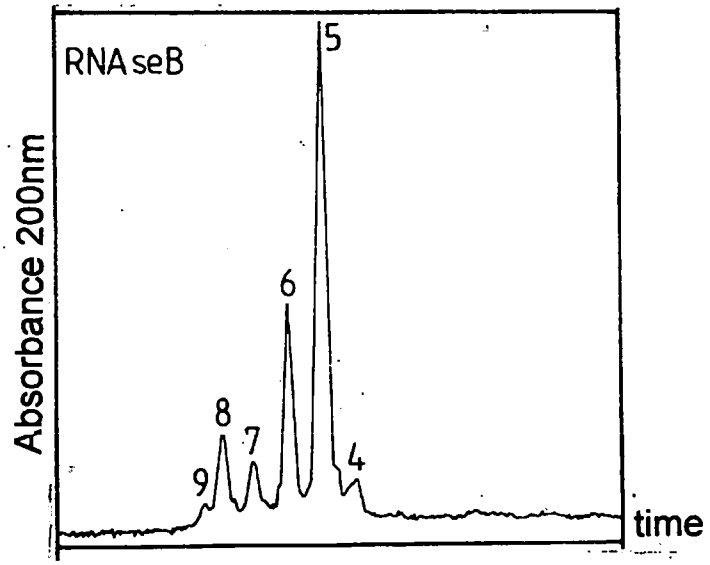


Figure 10: (a) The Man5-9 glycoforms of bovine pancreatic RNase B resolved in their correct molar proportions by CE. *A 72cm fused silica capillary, ID 75um, was used; voltage conditions were 1kv for 1min, 20kV for 9min, 300C, wavelength for detection 200nm, injection time 1.5 s, in 20mM sodium phosphate, 50 mM SDS, 5mM sodium tetraborate pH 7.2, on a Beckman P/ACE system.*

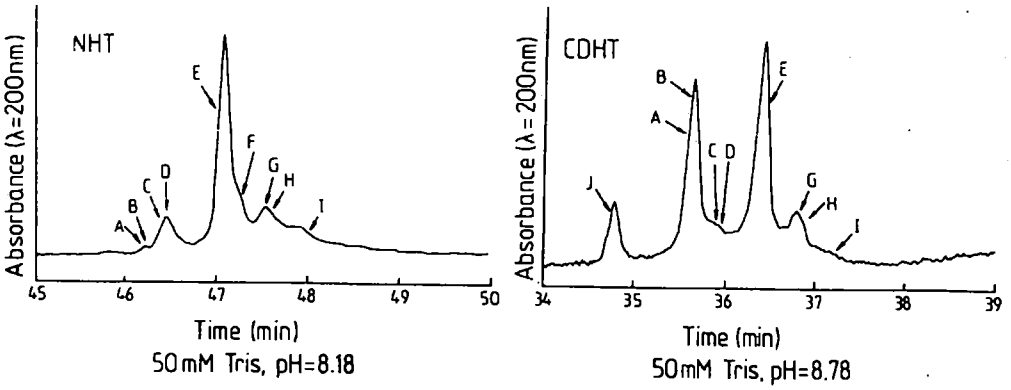


Figure 10: (b) The glycoforms of normal and carbohydrate deficient human transferrin resolved by CE. The structures A-J are shown diagramatically below in figure 10c.

Capillary Zone Electrophoresis **(CZE)** ***separations of transferrin were carried out on a P/ACE System 2100 (Beckman Instruments) using a fused silica capillary (107cmx75μm ID). The detection was by UV at 200nm. Injections of 1-2mg/ml for 1.5-2 seconds were performed (5ng) under high pressure. The buffer was 50mM Tris, 50mM sodium glutamate (CH_3CH_2COONa) (pH 8.54-8.64). The applied voltage was 10kV.***

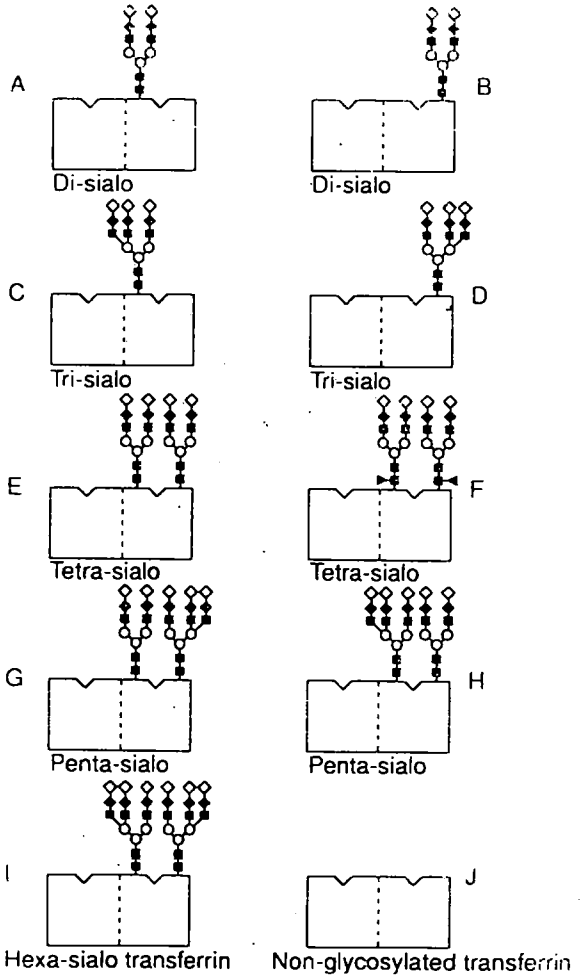


Figure 10c: Transferrin glycoforms present in CE analysis (Fig. 10b)

3. The glycosylation of neutrophil gelatinase B: a study using the technology which is currently available for the analysis of O-linked sugars.

3.1 Background:

Tissue plasminogen activator (tPA) contains three glycosyiation sites, one of which (Asn 184) is variably occupied. The enzyme occurs naturally in two forms: type I in which all three sites are occupied, and type II which contains only two N-linked oligosacccharides. Type II tPA is more active than type I towards its natural substrate, plasminogen (Parekh et al 1989). Plasminogen contains a variably occupied N-glycosylation site (Asn 297), and also occurs in two glycosylated forms: type 1 which contains one N and one O-linked glycan and type 2 which contains only the O-linked structure. In ch.7 it was established that the type of plasminogen modulates the kinetics of association of enzyme and substrate within the ternary complex of tPA, plasminogen and fibrin while the glycosyiation of tPA modulates the turnover rate. The combinatorial interaction of these variably glycosylated species allows a range of activities in which glycoforms of tPA and plasminogen which contain unoccupied glycosylation sites react faster than their fully glycosylated counterparts. This finding lead to the concept that glycosylation may be a means of control in the protease cascade reaction in which several glycosylated enzymes, including tPA, plasminogen and gelatinase and others, are involved. Matrix metalloproteinases, including collagenases, stromelysins and gelatinases, are key mediators in the physiological degradation of extracellular matrix molecules in tissue remodelling. They are also implicated in many pathologies including acute and chronic inflammation associated with diseases such as Rheumatoid Arthritis and Multiple Sclerosis (Opdenakker and van Damme 1994). Plasminogen activation is one of the activation processes of the matrix metalloproteinases, while activation of gelatinase A by stromelysin is another one. All these enzymes are interconnected in a cascade and gelatinase B is the most complex and terminal one. Three molecular types of gelatinase have been identified: (i) neutrophils stimulated with IL-8 secrete a 91kd gelatinase (calculated M.Wt. from amino acid sequence is 75kd) which is also found in high concentrations in the synovial fluid of arthritic joints (Opdenakker et al 1991, Masure et al 1991) (ii) normal human peripheral blood monocytes stimulated with IL-1 secrete an 85kd gelatinase (Opdenakker et al 1991) and (iii) the 92 kd and 85 kd forms have been characterised as truncated forms of the 96kd gelatinase from human tumor cells (Wilhelm et ai 1991, Van Ranst et ai 1991). Gelatinase B is the largest member of a family of structurally related matrix metalloproteinases. It contains seven protein domains, three potential N-glycosylation sites and a number of serine and threonine residues which may be O-glycosylated, although none has been reported. The experimental molecular weight of human neutrophil gelatinase B is 92 kD. Treatment with N-glycanase reduced this from 92kDa to 87kD (fig. 10d). There are no enzymes which cleave all O-linked glycans from proteins, therefore in this project hydrazinolysis and oligosaccharide analysis has been used to determine whether any of the additional 12kD of molecular weight can be attributed to O-glycosylation. in addition, the identification and analysis of the O-glycans has been

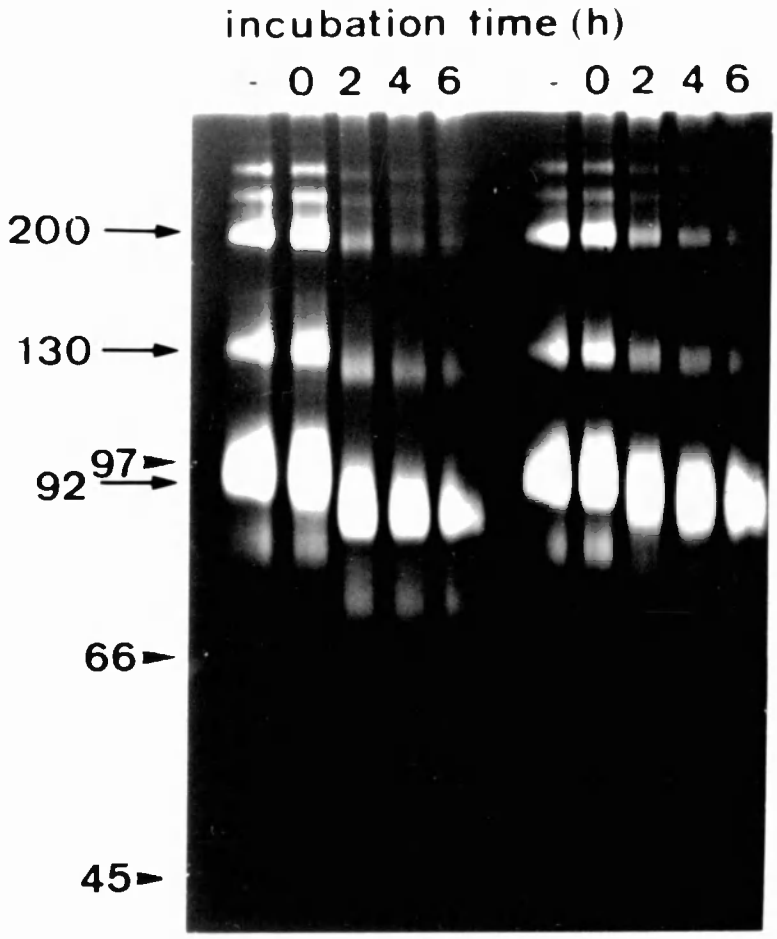


Figure 10d: Zymography of SDS PAGE of human neutrophil gelatinase B incubated with PNGase F

The native material migrates at 90-95kD, a homodimer migrates at approximately 200kD and a heterodimer with neutrophil gelatinase B associated lipocalin migrates at 130kD.

- : without enzyme; 0: with enzyme before incubation;

2, 4, 6: with enzyme and 2, 4 and 6 hours incubation respectively.

In tracks 1-5 the enzyme was used as recommended in the Oxford GlycoSystems Kit; tracks 6-10 used a 1:100 dilution of the enzyme supplied in the kit.

undertaken as a preliminary step to assess the possible role of O-glycosylation in the interaction of gelatinases with their natural substrates.

The project has also explored some of the limitations of the technology conventionally used to sequence O-glycans.

3.2 Methods

3.2.1 Preparation of gelatinase B

Gelatinase B was purified as described by Masure et al (1991). The 91kD form was analysed by amino terminal sequence analysis and was shown by zymography to be catalytically active.

3.2.2 Enzymatic deglycosylation of gelatinase B with N-glycanase

The purified enzyme was treated with PNGase F (N-glycanase) from different commercial sources (Oxford GlycoSystems -native form, and Boehringer-recombinant form) and analysed by zymography. This treatment reduced the molecular mass by approximately 5kDa to 87kDa.

3.2.3 Release and analysis of glycans

Glycans were released using the GlycoPrep 1000 optimised at 95°C for the release of N- and O-linked glycans. Reduction, radiolabelling and P4 GPC were carried out according to standard procedures described earlier in this chapter. Charge analysis used WAX chromatography and N-glycan analysis used the RAAM also described earlier. The analysis of O-linked glycans was by sequential exoglycosidase digestion.

3.3 Results: Analysis of charged glycans released from neutrophil gelatinase B

Gelatinase B was subjected to PNGase F to release N-linked glycans. Figure 10d shows the position of the native and deglycosylated material on an SDS gel, visualised by zymography. The native material migrates at 90-95kD (fig. 10d, tracks 1 and 6 labelled -). In addition a homodimer was identified at 200kD and a heterodimer with neutrophil gelatinase B-associated lipocalin (NGAL) also noted by Kjelsen et al (1993) was located at 120kD. The gradual disappearance of these additional bands after the removal of the N-linked sugars (fig. 10d: tracks 3-5 and 8-10; 2, 4, and 6 hours) may suggest a role for glycosylation in the formation of oligomers.

The sugars were also released from 40μg (10nmoles) of neutrophil gelatinase B (from Professor Ghislain Opdenakker; The Rega Institute, Leuven) using the OGS GlycoPrep 1000. (Yield specified by the manufacturer is >85%). The glycans were reduced and radiolabelled with tritium (total counts 1E8 cpm). 10% of the sample was subjected to charge analysis by WAX chromatography (Fig.11a). 30% of the structures were neutral. The remainder consisted of a mixture of sialylated N and O-links which were assigned by comparison with the analysis of fetuin sugars on the same system. After incubation with

neuraminidase and hydrolysis with acetic acid (1M 95^{0}C 1h) 90% of the sugars became neutral (Fig.11b). The remainder may be artefacts associated with radiolabelling but the possibility that they are sulphated or phosphated glycans has not been ruled out.

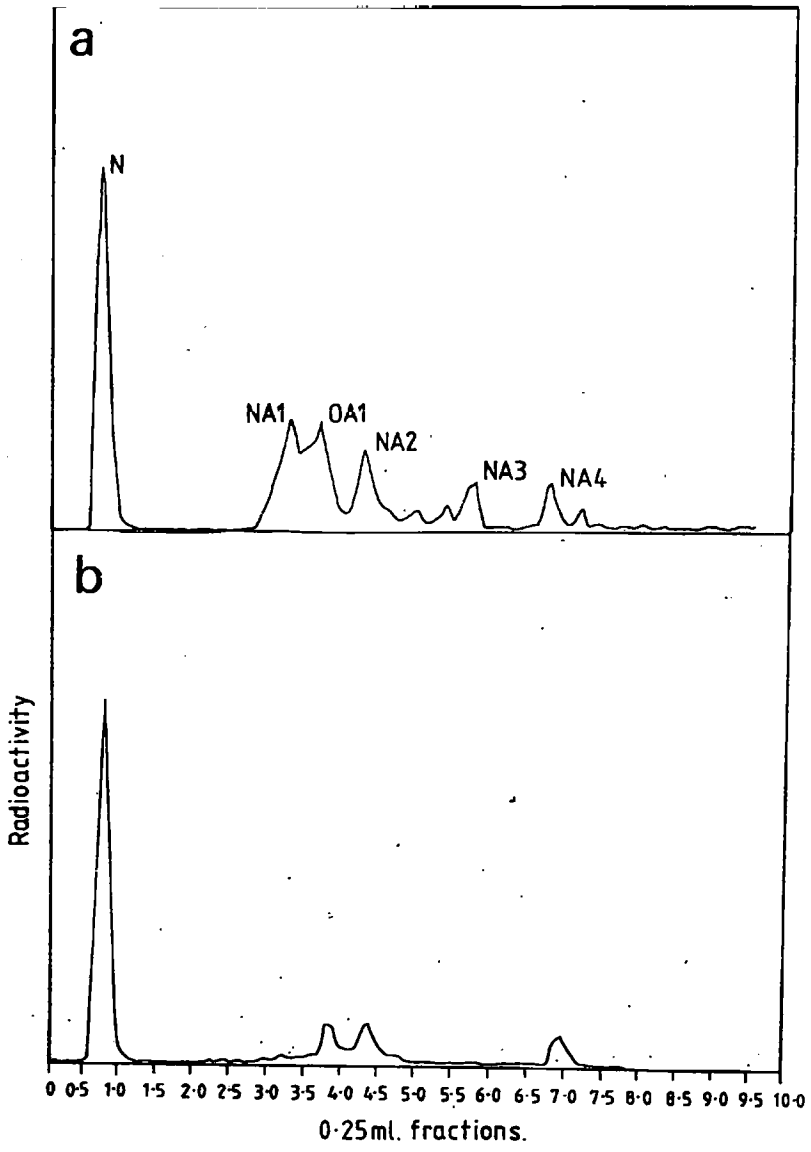


Figure 11: (a) Charge analysis of glycans released from neutrophil gelatinase B analysed by WAX chromatography ***(10E3 cpm was injected and the fractions were hand counted).*** **(b) Charge analysis of glycans released from neutrophil gelatinase B after incubation with** ***Arthrobacter neuraminidase*** **analysed by WAX chromatography.**

3.4 Analysis of the asialo glycans associated with neutrophil gelatinase

The released de-sialylated glycans were analysed by P4 GPC (Fig.12).

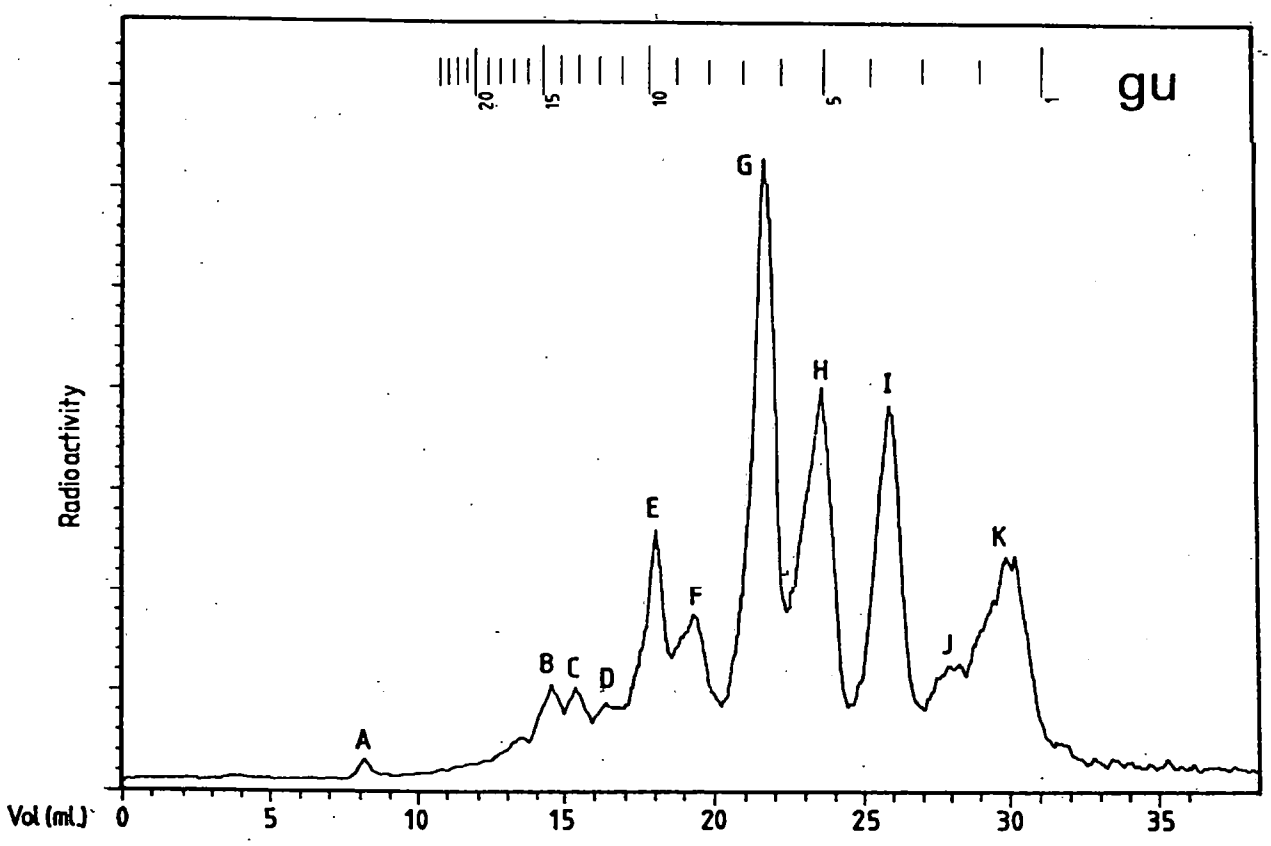


Figure 12: P4 GPC analysis of asialo glycans from neutrophil gelatinase.
Over 85% of the glycans eluted with molecular volumes typical of O-linked glycans (between 3.5 and 10.5gu), and 15% were larger structures (between 10.5 and 15.5gu).

Peaks B (15gu) and C (14.2 gu) (fig. 12) were rechromatographed and analysed by RAAM and identified as the N-linked glycans H5N4F2 and H5N4F respectively. Peaks E (9.7gu), G (6.5gu) and I (3.5 gu) were identified as O-linked glycans and analysed by sequential exoglycosidase digestions according to the scheme in figure 13. Peaks F (5.5 gu) and H (8.5

gu) were resistant to Jack Bean β–hexosaminidase and Jack Bean α–mannosidase and remain to be assigned. It is possible that they may contain α-linked glycans or fucose at the terminal non reducing positions. The void (A) contained a small amount of non-carbohydrate material, and peak D contained too little material to be analysed. J and K at 2.5 and 1.5 gu respectively were not analysed, they contain material of very low molecular volume and, although the possibility that they are derived from gelatinase cannot be ruled out, these elution positions are generally associated with environmental contaminants.

3.5 Sequential exoglycosidase analysis of O-linked glycans

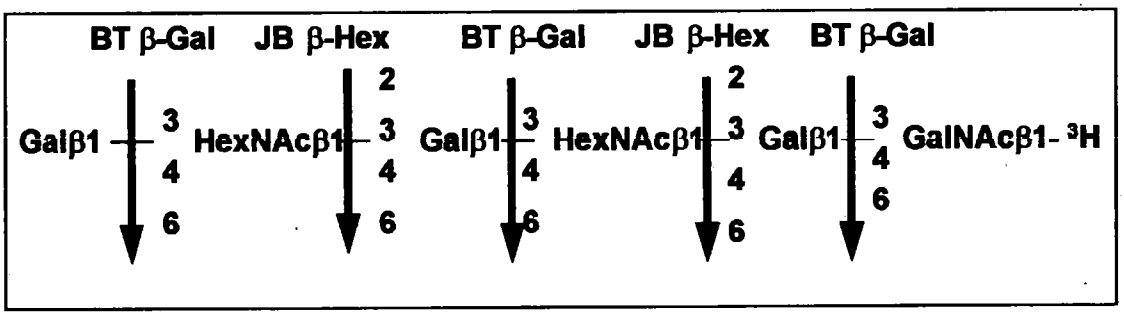


Figure 13: Scheme for sequential exoglycosidase sequencing of O-linked glycans

BT β-gal: Bovine testes β- galactosidase;

JB β-hex: Jack bean β-hexosaminidase

Peak G (6.5gu) was re-chromatographed (Fig. 14a) and the central peak fractions pooled to obtain a relatively pure sample of the major glycan in the pool. The aqueous solution of sugars was evaporated to dryness and incubated with Bovine testes β-galactosidase. The digest was purified on mixed bed ion exchange resins and chromatographed on P4 where it eluted at 5.65gu (Fig. 14b). This indicated that the glycan had been reduced by one glucose unit and that the terminal sugar which had been removed was 1-3,4 or 6 linked galactose. The oligosaccharide eluted from the P4 column was pooled, dried, and incubated with Jack Bean β–hexosaminidase. On re-chromatographing the glycan eluted at 3.67gu (14c) indicating that 1-2,3,4 or 6 linked N-acetylglucosamine or N-acetylgalactosamine had been removed from the non-reducing terminus. The glycan was again digested with bovine testes β-galactosidase to a glycan eluting at 2.69gu (14d) indicating that the terminal residue was 1-3,4 or 6 linked galactose. Finally the 2.69gu peak was analysed by gas chromatography which showed that the reducing terminal radiolabelled monosaccharide was N-acetylgalactosamine (14e).

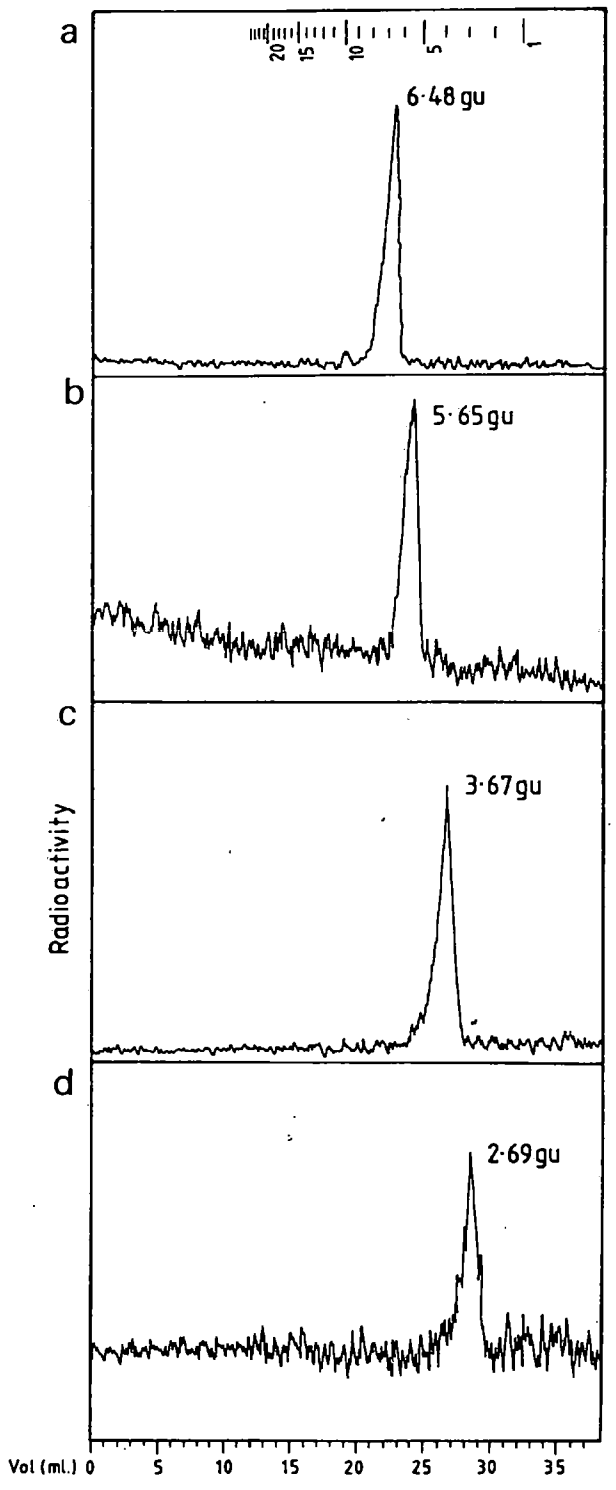


Figure 14a-d: Exoglycosidase sequencing of peak G

Peaks E and I were sequenced in an analogous manner and it was finally established that in figure12 peaks E,G and I were members of a series of O-linked glycans:

Pk E:	Galβ1-(4,3,6)HexNAcβ1-(2,3,4,6)Galβ1-(4,3,6)HexNAcβ1-(2,4,3,6)Galβ1-(3)(4,6) GalNAc
Pk G:	Galβ1-(4,3,6)HexNAcβ1-(2,4,3,6)Galβ1-(3)(4,6) GalNAc
Pk I:	Galβ1-(3)(4,6) GalNAc

Table 3: The series of O-linked glycans associated with gelatinase B. *The relative proportions of the glycans in these peaks (E:G:I) was 1:2.6:1.5*

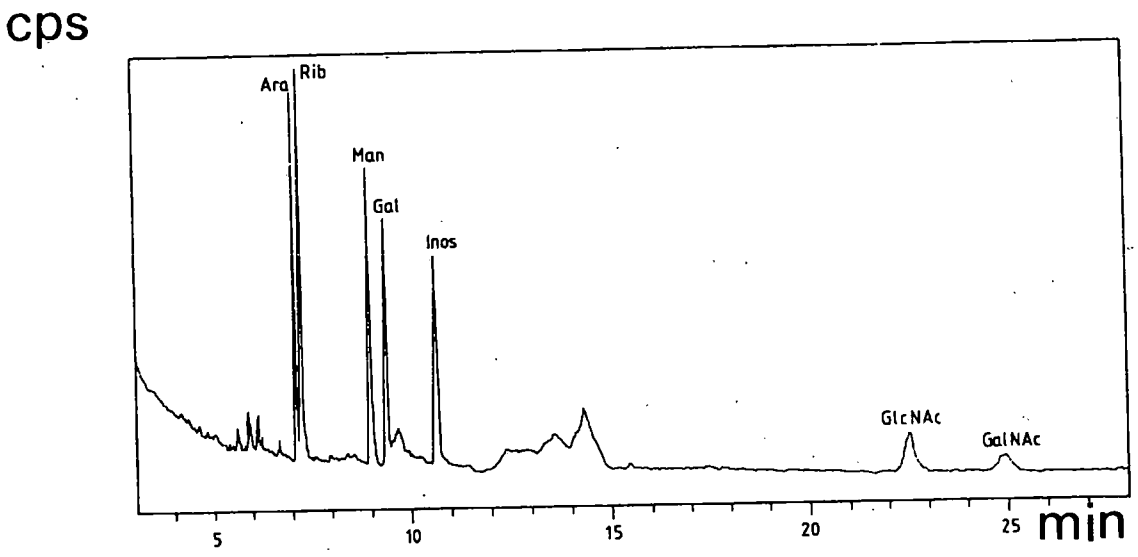


Figure 14e: RadioGC analysis of terminal monosaccharide from peak G. *The standard monosaccharides Ara, Rib, Man, Gal, Inos and GlcNac were co-injected with the sample which was identified as GalNAc.*

3.5 Discussion

Sequencing O-linked glycans

This project highlights the need to develop new technology to automate the sequencing of O-linked glycans. The existing RAAM technology is unsuitable for analysing repeating subunits, and currently such structures are determined by repetitive digestions.

In addition, more enzymes are required to determine the structure of the glycosidic linkages. For example, Jack Bean β hexosaminidase is relatively non-specific and digests either GlcNAc or GalNAc linked β1-2,3,4 or 6. The only alternative enzymes currently available are highly specific and therefore cannot be used when there are limited amounts of material. For example chicken liver β-hexosaminidase will digest only GlcNAc β1-3,4. Such specific enzymes are essential for this type of sequencing, but when supplies of material are limited, as is the case with gelatinase B, enzymes of intermediate selectivity may also be required. O-linked sugars are not normally suitable for MS analysis because they have molecular weights in the same range as the matrix components, although labelling with 2AB (120d) can in some cases overcome this problem.

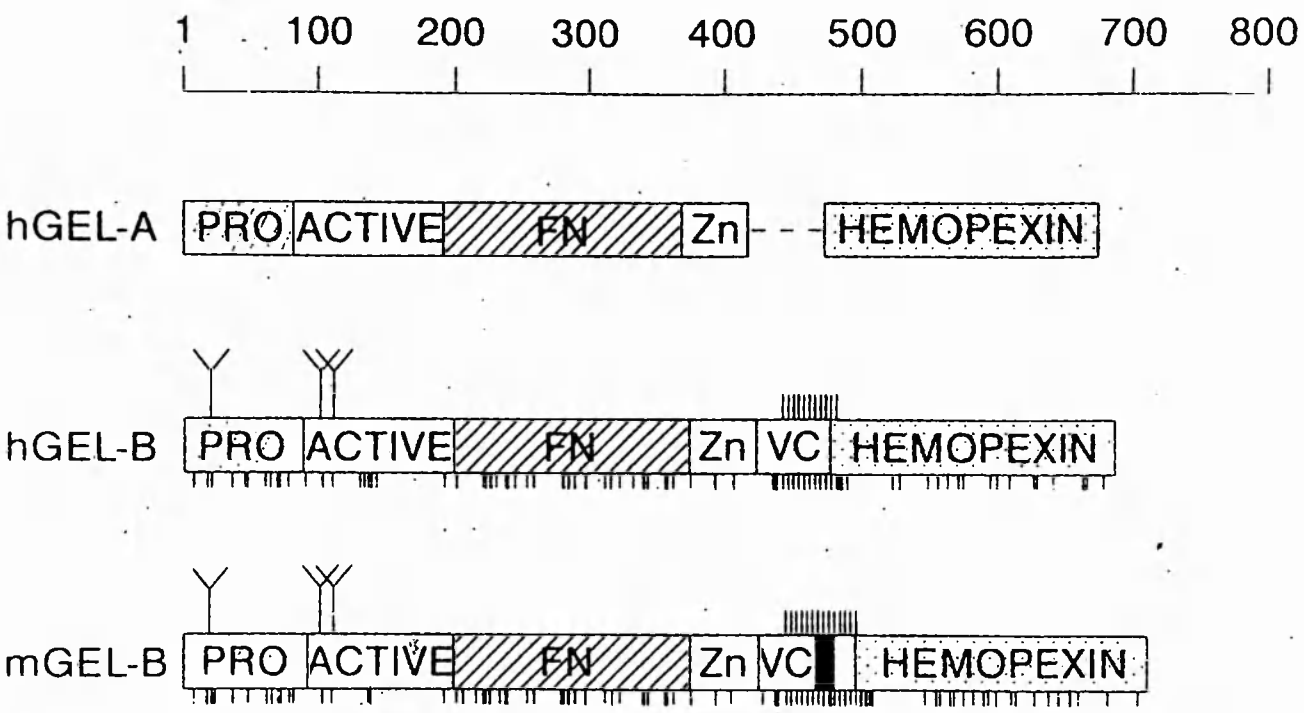


Figure 15: Diagram of domains in the proenzymes human gelatinase A and B (hGEL-A and hGEL-B) and murine gelatinase B (mGEL-B)

The scale at the top shows the number of amino acids in the protein relative to the domains in hGEL-A and -B and mGEL-B. The enzymes are built from 6 or 7 domains some of which are homologous with previously characterised domains in other proteins: e.g. FN: fibronectin; Zn: Zinc binding domain; VC: collagen type V domain and hemopexin.

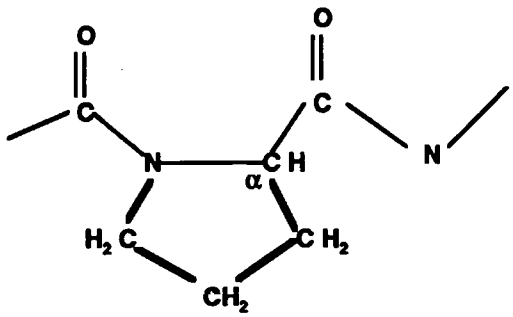
N-linked glycosylation sites are denoted by the symbol Y, serine and threonine residues which are potential sites for O-glycosylation are denoted by the symbol I. Murine and human gelatinase B contain a section of primary sequence rich in proline, serine and threonine, which is not conserved in human gleatinase A.

The glycosylation of gelatinase B from human neutrophils

The analysis indicates that gelatinase B contains a series of glycans with a type 1 core structure (Galβ1,3GalNAcR). These have not been reported previously, and contribute to the observed additional molecular weight of the enzyme which is not accounted for by the amino acid composition and the N-glycans. The proportion of N-linked glycans is relatively low in proportion to the O-links, and may be even less than these data suggest since, as discussed earlier, the conditions used to release and recover the glycans allow a higher proportion of the total N-links to be recovered compared with the O-links. Two N-linked structures were identified, H5N4F2 and H5N4F. Core type I O-glycans are synthesised by stepwise addition of monosaccharide residues to the GalNAc core, beginning with β1-3 GalNAc. Elongation can proceed along a number of pathways, the most common is the addition of Galβ1-4GlcNAc, and the results of this analysis are consistent with this. A common feature of core I type O-links is the presence of α-linked terminal residues, such as α2-3 or 6 linked sialic acid and fucose, at the reducing terminus. Consistent with this WAX chromatography (Fig. 11) indicated that gelatinase contains mono-sialylated 2,3 or 2,6 linked sugars which can be removed by incubation with *Arthrobacter ureafaciens* neuraminidase.

Structure and function of O-linked oligosaccharides attached to gelatinase

The amino acid sequence of human and mouse gelatinase B as deduced from molecular cloning experiments (Wilhelm et al 1989, Masure et al 1993) indicates that the molecule contains many serine and threonine residues a number of which are in the vicinity of proline residues. In particular, in both murine and human 92kD gelatinase B there is a section of the primary sequence between residues (in human residues 452 to 503) which is rich in Pro, Ser and Thr residues (fig. 15). Such a region of amino acids is characteristic of a peptide region which may contain a cluster of O-linked glycans. Interestingly this region, which is homologous with a portion of the helical region of α2(V) collagen (Weil et al 1987), is not present in other members of the secreted metalloprotease gene family. Most O-linked glycans are attached to the hydroxyl groups in the side chains of either threonine or, less often, to serine residues through GalNAc, although glycans such as GlcNAc (Hart et al 1988) or fucose (Harris et al 1991) may also be O-linked to the peptide backbone. There is no consensus sequence in the peptide for GalNAc O-linked to Thr or Ser, however it is common to find proline in close proximity. This may be because the nitrogen in the backbone is part of the proline side chain and limiting the conformations available to the protein (Branden and Tooze 1991). This prevents the N atom from hydrogen bonding and also provides some hindrance to the α-helical conformation. Although it fits well into the first turn of an α-helix, in other positions proline normally forms a bend in the peptide chain which exposes the residues which are close to proline in the primary amino acid sequence.



Proline

Molecules which contain a significant proportion of O-linked glycans, such as mucins and CD5 (ch2), frequently contain stretches of peptide sequence with repeating serine, threonine and proline residues. The clustered O-linked glycans repel each other, particularly when sialylated, and their structural role may be to extend the polypeptide chain. Sections of the peptide which contain many proline residues in the chain assume a helical structure. When sugars are attached to serine or threonine within these helices the protein and sugars together may assume a 'bottle brush-like' structure. O-linked sugars may also inhibit proteolytic digestion of peptides and contribute to protein stability. The activity of human granulocyte colony stimulating factor is largely eliminated on removal of the O-linked sugars and it has been proposed that the sugars maintain the protein conformation and prevent polymerisation (Oh-eda et al 1990). O-linked sugars may also be involved in binding events, for example on the zona pellicuda of mouse eggs O-linked sugars are implicated in the binding of the glycoprotein ZP3 to the sperm receptor (Florman and Wassarman 1985) and O-linked glycans on IgA bind to a T-cell receptor which is upregulated in rheumatoid arthritis (Rudd et al 1994). O-linked glycans, including blood group antigens, often show significant alterations in diseases such as cancer and in inflammation. Neutrophil gelatinase B is found in high concentrations in the synovial fluid of rheumatoid patients. However, the implications of the O-glycosylation of gelatinase for the structure and function of the enzyme have not yet been assessed; this is a preliminary step in the investigation of the role of glycosylation in an enzyme which, in common with tPA, and plasminogen is involved in the re-modelling of the extra-cellular matrix.

References to chapter 8:

Ashford, D., Dwek, R.A., Welply, J.K., Amatayukul, S., Homans, S.W., Lis, H., Taylor, G.N., Sharon, N. and Rademacher, T.W. (1987) Eur. J. Biochem. 166 311-320 The β1-2-D-xylose and α1-3-L fucose substituted N-linked oligosaccharides from *Erythrina Crystagalli* lectin. Isolation, characterisation and comparison with other legume lectins.

Beavis, R.C. and Chaitt, B.T., (1990) Proc. Natl. Acad. Sci. USA 87 6873-6877 Rapid, sensitive analysis of protein mixtures by mass spectrometry.

Beavis, R.C., Chaudhary, T. and Chaitt, B.T. (1992) Org. Mass Spectrom. 27 156 Alpha-cyano-4-hydroxy-cinnamic acid as a matrix for MALDI MS.

Brandon, C. and Tooze, J. (1991) In : Introduction to protein structure (ch.1: Basic structural principles) Garland Publishing Inc. N.Y. and London.

Dwek, R.A., Edge, C.J., Harvey, D.A. and Wormald, M.W. and Parekh, R.B. (1993) Ann. Rev. Biochem. 62 65-100 Analysis of glycoprotein associated oligosaccharides.

Edge, C. J., Rademacher, T.W., Wormald, M., Parekh, R.B., Butters, T.D., Wing, D.R. and Dwek, R.A. (1992) PNAS (USA) 89 6338-42 Fast sequencing of oligosaccharides: the reagent array analysis method.

Florman, H.M. and Wassarman, P.M. (1985) Cell 41 313-324 O-linked oligosaccharides of mouse egg ZP3 account for its sperm receptor activity.

Guile, G.R., Wong, S.Y.C. and Dwek, R.A. (1994) Analytical Biochemistry 222 231-235 Analytical and preparative separation of anionic oligosaccharides by weak anion exchange high performance liquid chromatography on an inert polymer column.

Harris, R.J., Leonard, C.K., Guzetta, A.W., and Spellman, M.W. (1991) *Biochemistry* 30, 2311-2314 Tissue plasminogen activator has an O-linked Fucose attached to Threonine-61 in the Epidermal Growth Factor Domain.

Hart, G.W., Holt. G. and Haltiwanger, R.S. (1988) TIBS 13 380-384 Nuclear and cytoplasmic glycosylation: novel saccharide linkages in unexpected places.

Harvey, D.J. and Rudd, P.M; Bateman, R.H., Bordoli, R.S., Howes, K., Hoyes, J.B. and Vickers, R.G. (1994) Organic Mass Spectrometry 29 753-765 Examination of complex oligosaccharides by matrix -assisted laser desorption mass spectrometry on time-of-flight instrument and magnetic sector instruments.

Hillenkamp, F. and Karas, M. (1990) Methods in Enzymology 193 280-295 Mass spectrometry of peptides and proteins by Matrix Assisted Ultra Violet Laser Desorption/ Ionisation.

Jackson, P. Anal. Biochem. (1994) 216, 243-252 The analysis of Fluorophore-labelled glycans by high-resolution polyacrylamide gel electrophoresis.

Jorgensen, J.W. and Lukacs, K.D. (1983) Science 222 266-268 Capillary Electrophoresis

Karas, M. and Hillenkamp, F.(1988) Anal. Chem. 60 2299-2301 Laser desorption ionisation of particles with molecular weights exceeding 100,000 daltons.

Kjeldsen, L., johnsen, A.H., Sengalo, H. and Borregaard, N. (1993) Isolation and primary structure of NGAL, a novel protein asssociated with human neutrophil gelatinase.

Masure, S., Proost, P., Van Damme, J. and Opdenakker, G. (1991) Eur. J. Biochem. 198 391-398 Purification and identification of 91-kD neutrophil gelatinase. Release by the activating peptide interleukin-8.

Mori, K., Dwek, R.A., Downing, A.K., Opdenakker, G., and Rudd, P.M. (1995) J. Biol. Chem. 270 3261-3267 The activation of type 1 and type 2 plasminogen by type I and type II tissue plasminogen activator.

Oh-eda, M., Hasagawa, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T., Tomonou, K., Yamaziki, T. and Ochi, N. (1990) J. Biol. Chem. 265 11432-11435 O-linked sugar chain of human granulocyte colony stimulating factor protects it against polymerisation and de-naturation allowing it to retain its biological activity.

Opdenakker, G., Masure, S., Proost, P., Billiau, A. and Van Damme, J. (1991) FEBS Letters 284 73-78 Normal human monocyte gelatinase and its inhibitor.

Opdenakker, G. and van Damme, J. (1994) Immunology Today 15 103-107 Cytokine regulated proteases in autoimmune diseases.

Parekh, R.B., Dwek, R.A., Thomas, J.R., Opdenakker, G., Rademacher, T., Wittwer, A.J., Howard, S.C., Nelson, N., Siegel N.R., Jennings, M.G., Harakas, N.K. and Feder, J. (1989) Biochemistry 28 7644-7662 Cell-Type Specific and Site Specific N-Glycosylation of Type I and Type II Human Tissue Plasminogen Activator.

Patel, T., Bruce, J., Merry, A., Bigge, C., Wormald, M., Jaques, A. and Parekh, R.B. (1993) Biochemistry 32 679-693 Use of hydrazine to release in intact and unreduced form both N- and O-linked oligosaccharides from glycoproteins.

Prime, S. and Dearnley, J. (1993) Fast Oligosaccharide Sequencing Using the Reagent Array Analysis. Oxford GlycoSystems Application note.

Rudd, P.M., Fortune, F.M., Patel, T., Parekh, R.B., Dwek, R.A. and Lehner, T. (1994) Immunology 83 99-106 IgA Binding to T-cell Surface Receptor involves "O"-linked Sugars from the Hinge Region of IgA1.

Rudd, P.M., Scragg, I.G., Coghill, E. and Dwek, R.A. (1992) Glycoconjugate Journal 9 86-91 The Separation and Analysis of the Glycoform Populations of Ribonuclease B using Capillary Electrophoresis.

Rudd, P.M., Mattu, T. and Honda, S. (1995 *in press*) in 'Capillary Electrophoresis: A Practical Approach' (OUP) Ed. D. Goodall. Capillary Electrophoresis of Oligosaccharides, Glycoproteins and Glycopeptides.

Stahl, B., Steup, M., Karas, M. and Hillenkamp, F. (1991)Anal. Chem. 63 1463 Analysis of neutral oligosaccharides by MALDI MS.

Van Ranst, M., Norga, K., Masure, S., Proost, P., Vanderkerckhove, J., Auwerx, J. Van Damme, J. and Opdenakker, G. (1991) Cytokine 3 231-239 The cytokine-protease connection: identification of a 96kD THP-1 gelatinase and regulation by interleukin-1 and cytokine inducers.

Wilhelm, S.M., Collier, I.E., Marmer, B.L., Eisen, A.Z., Grant G.A. and Goldberg, G.I. (1989) J. Biol. Chem. 264 17213-17211 SV40-transformed human lung fibroblasts secrete a 92-kDa type IV collagenase which is identical to that secreted by normal human macrophages.

Weil, D., Bernard, M., Gargano, S., Ramirez, F. (1987) Nucleic Acids Res. 15 182-198